Stochastic Optimization under Probust and Dynamic Probabilistic Constraints: with Applications to Energy Management

Dissertation
zur Erlangung des akademischen Grades

doctor rerum naturalium
Dr. rer. nat.
im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt- Universität zu Berlin

von
Frau M.Sc. Tatiana Carolina González Grandón

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I:
Prof. Dr. Elmar Kulke

Gutachter

Prof. Dr. Werner Römisch
Prof. Dr. Rüdiger Schultz
P.D. Dr. René Henrion

Tag der mündlichen Prüfung 31. Mai 2019

Ich erkläre, dass ich die Dissertation selbstständig und nur unter Verwendung der von mir gemäß § 7 Abs. 3 der Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fakultät, veröffentlicht im Amtlichen Mitteilungsblatt der Humboldt-Universität zu Berlin Nr. 126/2014 am 18.11.2014 angegebenen Hilfsmittel angefertigt habe.

Berlin, den 26. Februar 2019

Abstract

Optimization problems under uncertain conditions abound in many energy management applications. Once the relevant uncertainties are identified for the optimization model, there are two major classes of approaches to employ: the stochastic programming and the robust optimization methods. This thesis offers, in the first and second chapter, a general overview of the classical approaches to solving optimization under uncertainty, with a focus on probabilistic constraints.

Then, in the third chapter, a new class of so-called **Probust constraints** is introduced in the presence of models with uncertain parameters having partially stochastic and partially non-stochastic character. We show the relevance of this class of approach and solve two problems in a stationary gas network. First, in the context of gas transportation, one ends up with a constraint, which is probabilistic with respect to the load of gas and robust with respect to the roughness coefficients of the pipes (which are uncertain due to a lack of attainable measurements). Secondly, we solve a problem for a network operator, who would like to maximize the offered capacity for old and new customers. In this case, one is faced with an uncertain total demand which is probabilistic for old clients and robust for new clients. In both problems, we demonstrate how probust constraints can be dealt within the framework of the so-called spheric-radial decomposition of multivariate Gaussian distributions.

Furthermore, in chapter four, we present novel structural and numerical results for optimization problems under a **dynamic joint probabilistic constraint**. This constraint is of the type

$$\mathbb{P}(g_i(x_1, x_2(\xi_1), x_3(\xi_1, \xi_2) \ldots, x_T(\xi_1, \ldots, \xi_{T-1}), \xi_1, \ldots, \xi_T) \leq 0 \quad (i = 1, \ldots, m)) \geq p, \tag{1}$$

where $(\xi_1, \ldots, \xi_{T-1})$ is a finite stochastic process, $(x_1, \ldots, x_T)$ is an adapted process of decision policies depending on previously observed outcomes of the random process, $\mathbb{P}$ is a probability measure and $p \in [0, 1]$ is a probability level. Strong and weak semicontinuity results are obtained for the general case depending on whether policies are supposed to be in L^p or $W^{1,p}$ spaces. For a simple two-stage model, verifiable conditions for Lipschitz continuity and differentiability of this probability function are derived and endowed with explicit derivative formulae. Numerical results are illustrated for the solution of a two-stage hydro-power reservoir where (1) figures as a constraint.

Zusammenfassung

Optimierungsprobleme unter Unsicherheiten treten in vielen Anwendungen des Energiemanagements auf. Wenn die relevanten Unsicherheiten für das Optimierungsmodell identifiziert sind, gibt es zwei Hauptklassen von Ansätzen: die stochastische Programmierung und die robusten Optimierungsmethoden. Diese Arbeit liefert, in den ersten beiden Kapiteln einen allgemeinen Überblick über die klassischen Ansätze zur Optimierung unter Unsicherheit mit einem Schwerpunkt auf probabilistischen Randbedingung.

Anschließend wird im dritten Kapitel eine neue Klasse von sogenannten **Probust Randbedingungen** beim Auftreten von Modellen mit unsicheren Parametern mit teilweise stochastischem und teilweise nicht-stochastischem Charakter eingeführt. Wir zeigen dabei die Relevanz dieser Aufgabentypen für zwei Problemstellungen in einem stationären Gasnetz auf. Erstens liegen beim Gastransport probabilistische Randbedingungen bezüglich der Gasnachfrage vor sowie auch robuste Randbedingungen bezüglich der Rauheitskoeffizienten in den Rohren, welche in der Regel unbekannt sind, da es keine zuverlässigen Messmöglichkeiten gibt. Zweitens lösen wir ein Problem für einen Netzbetreiber, der zum Ziel hat, die angebotene Kapazität für alte und neue Kunden zu maximieren. In diesem Fall ist man mit einer ungewissen Gesamtnachfrage konfrontiert, die sich aus der probabilistischen Nachfrage für Altkunden und der robusten Nachfrage für Neukunden zusammensetzt. Für beide Fälle zeigen wir, wie mit probusten Randbedingungen im Rahmen der sogenannten sphärisch-radialen Zerlegung multivariater Gauß-Verteilungen umgegangen werden kann.

Des Weiteren präsentieren wir in Kapitel vier neue strukturelle und numerische Ergebnisse für Optimierungsprobleme unter einer **dynamischen gemeinsamen probabilistischen Randbedingung**. Diese Randbedingung ist vom Typ

$$\mathbb{P}(g_i(x_1, x_2(\xi_1), x_3(\xi_1, \xi_2) \ldots, x_T(\xi_1, \ldots, \xi_{T-1}), \xi_1, \ldots, \xi_T) \leq 0 \quad (i = 1, \ldots, m)) \geq p,$$

wobei $(\xi_1, \ldots, \xi_{T-1})$ ein endlicher stochastischer Prozess ist, $(x_1, \ldots, x_T)$ ein angepasster Prozess von Entscheidungsregeln in Abhängigkeit von zuvor beobachteten Ergebnissen des Zufallsprozesses, $\mathbb{P}$ ein Wahrscheinlichkeitsmaß und $p \in [0, 1]$ ein Wahrscheinlichkeitsniveau. Starke und schwache Halbstetigkeitsergebnisse werden für den allgemeinen Fall, in Abhängigkeit davon ob Strategien in L^p oder $W^{1,p}$ Räumen angenommen werden, erstellt. Für ein einfaches zweistufiges Modell werden überprüfbare Bedingungen für die Lipschitz-Stetigkeit und die Differenzierbarkeit dieser Wahrscheinlichkeitsfunktion abgeleitet und mit expliziten Ableitungsformeln unterstützt. Numerische Ergebnisse werden für die Lösung einer zweistufigen Steuerung eines Wasserkraftwerkspeichers veranschaulicht wobei (1) als Begrenzung gilt.

A mis adorados mami y papi, que nacieron con corazones guerreros,
y la vida les ha hecho luchar sin cesar.

Acknowledgements

First of all, I would like to express my deepest appreciation and gratitude to my supervisor PD Dr. René Henrion for his excellent scientific guidance through the past three years. From the beginning, he provided me with a very interesting and challenging research project on new topics dealing with probabilistic constraints. I have benefited intellectually from his profound knowledge in stochastic optimization; nonsmooth analysis and his skillful ideas on applications. Besides, I have also enjoyed a myriad of fruitful discussions on historical and political views. I could not have imagined having a better advisor for my Ph.D. study.

I would also like to extend my sincere thanks to Prof. Dr. Alejandro Jofre for organizing the stay at the *Center for Mathematical Modelling (CMM)* at Universidad de Chile. During this research stay, he provided me with an interesting problem and I was able to deepen my knowledge of optimization in Banach spaces. I would also like to thank my colleague and friend Dr. Pedro Pérez-Aros for an interesting cooperation and readiness to discuss mathematical problems. This thesis has strongly benefited from inspiring discussions and ideas obtained during this research visit.

I owe many thanks to the members of my jury committee, Prof. Dr. Caren Tischendorf, Prof. Dr. Werner Römisch and Prof. Dr. Rüdiger Schultz, for their time and valuable comments. I want to thank Prof. Dr. Ulrich Horst for being my mentor during the five years of *BMS* graduate program and Prof. Dr. Franz Hubert for the opportunity to research and teach in his Chair on the last semester of my Ph.D. thesis.

Special thanks to Dr. Holger Heitsch for our collaboration, and stimulating discussions related to probust constraints and numerical results in the context of gas networks. My warm thanks go to my colleagues and friends Dr. Johann Schmitt and Jakub Marian for reading some chapters of the thesis and contributing with constructive comments. In particular, I thank Felipe Serrano for introducing me to an advanced use of MATLAB.

In addition, I would like to thank Dr. Forough Sodoudi, the managing director of the *Berlin Mathematical School (BMS)* for her invaluable endorsement and guidance in the transit from Phase-I to Phase-II of the BMS graduate program.

I am very grateful to the *BMS* and *TRR 154 Graduate College* for the opportunity of studying mathematics in an advanced level as well as the *DFG, CONACYT, TRR154 Graduate College*, and the *Stiftung der Deutschen Wirtschaft (SDW)* for their financial support in different stages of the five years of my graduate program.

Last but not least, I would like to express my admiration and heartfelt gratitude to my future husband Vivien Barnier for his unrestrained love, support and emotional backup through everyday of my Ph.D. program. We made it together.

Contents

Introduction

The introduction of mathematical optimization techniques in the 1950s and 1960s, together with the commercial availability of digital computers, had a strong influence on energy management. In the past, energy companies were vertically integrated, monopolistic and followed a centralized production planning mode; they used mathematical optimization tools mostly focused on engineering analysis. Linear and non-linear programming had applications in resource allocation, refinery, distribution, transportation of oil (see e.g. [127, 54, 89, 110]). Also dynamic programming, pioneered by Richard Bellman [11], was a successful procedure for multi-stage optimization problems, which had widespread use, for instance, in gas field optimization and pipeline design [147]. At this time, energy management was technical in nature.

However, the regulatory framework of energy markets changed towards a liberalization paradigm at the end of the 20th century. In this regard, Chile (1982) and the United Kingdom (1989) were the pioneer countries experiencing energy market liberalization, privatization and restructuring of the energy supply and distribution industry. The creation of an electricity and gas market in Europe started in 1995. This change of structure of ownership and regulation changed the way decision making in energy management is achieved. Moreover, the Kyoto Protocol in 2005 also enhanced the promotion of renewable energies and set climate policy on the global agenda. Today, stakeholders in the energy sector face problems of technical, economic and environmental nature. This had two consequences for optimization models in the energy management sector.

First, there are different types of economic configurations managing energy systems: centrally planned, deregulated or public-private partnership. According to each of these structures, the aim of the energy company changes. The objective of public companies is to minimize costs of delivery to consumers; policy makers maximize social welfare to understand market distortions; while private companies strive to maximizing profits. Secondly, the amount of input data and restrictions is not only much larger, but many parameters on the constraints are uncertain and volatile. New markets have emerged together with the electricity and gas, like CO_2

emission rights and financial derivative markets. Forecasts of fuel prices and intermittent renewable energy generation are by far not the only sources of uncertainty and therefore of commercial risks. There is uncertainty in hourly spot prices, demand of energy, heat load forecasts, technical failures due to uncertain weather conditions and risks of not satisfying the demand-offer balance. Thus, it is undeniable that the sources of uncertainties and the amount of historical data available are increasing in decision making process of energy management. Moreover, uncertainty about input parameters leads to imbalances and wrong management decisions [126]. In the future, energy markets will become even more volatile and complex due to the large-scale shift to intermittent renewable technologies, resource scarcity, energy sector coupling, CO_2 quotas and the advent of supergrids. So, how can energy decision makers handle uncertainty?

Since 1990, two major classes of methods of optimization problems under uncertainty emerged in energy decision making: the stochastic optimization and the robust optimization methods. In the stochastic optimization approach, uncertain inputs are modeled as continuous or discrete random variables with known probability distribution. In contrast, the robust optimization approach does not consider the probability distribution of uncertain input variables; rather these input variables are assumed to take values within some predefined uncertain interval or set. It is well known that the stochastic optimization method has two major classes: the recourse approach and the probabilistic constrained optimization approach. In recourse-based optimization, decisions are taken prior to the realization of uncertain parameters. Then, after the realization of uncertain parameters, recourse or corrective measures are taken to compensate for possible violation of constraints. In many applications, however, compensations simply do not exist (ecological damages, engineering safety, flood restrictions of a water reservoir) or the compensations cannot be quantified. In such applications, one would rather take decisions that guarantee feasibility with a high probability.

In this thesis, we first present an overview of the mathematical approaches to deal with optimization under uncertainty with continuous random variables. We mainly focus on probabilistic constraints. Such constraints allow one to find optimal decisions, which are probabilistically robust against uncertainty. The main contributions of the thesis, exposed in ch. 3 and ch. 4, address the theoretical analysis and numerical treatment of two brand-new topics in connection with probabilistic constraints: the family of probabilistic/robust (*probust*) constraints, where the joint presence of uncertainty in stochastic and unspecified form can be handled in a single model; and optimization under dynamic probabilistic constraints, which take

into account in their decision policies the flow of information, while randomness unravels. Hence the thesis is organized as follows.

Structure of the Thesis

Chapter 1 gives an introduction to four different ways of incorporating uncertainty into an optimization model. Expected value model, recourse based model, probabilistic constrained optimization and robust optimization are discussed. Probabilistic constrained optimization is explained in detail, since it is the backbone of our main research problems.

Chapter 2 gives a brief introduction to an efficient method to solve probabilistic constrained optimization problems via numerical nonlinear optimization for the case of normally distributed random parameters. One can calculate the functional values of the probabilistic constraint and its gradients via the spherical-radial decomposition method. With this method, it is possible to solve corresponding optimization problems for Gaussian and Gaussian-like random vectors in dimension of up to a few hundred.

Chapter 3 introduces a new class of joint probabilistic/robust constraints, so-called *Probust Constraints*. A convenient combination with ideas from robust optimization then leads to probust functions, i.e., probability functions acting on generalized semi-infinite inequality systems. In this way, the joint presence of uncertainty in stochastic and unknown form can be handled in a single model. Relevance of the model, as well as new algorithms, is illustrated through two examples in the gas sector: gas transport and gas expansion capacity. For both models, one is able to describe a methodology for the solution as well as a numerical solution based on the spherical-radial decomposition. The content of this chapter is published in [60] and [4].

Finally, in **Chapter 4** we move to an infinite-dimensional setting and elaborate on the structural and numerical analysis of *Dynamic Joint Probabilistic Constraints*. This model takes into account in their decision policies the flow of information while randomness unravels. The use of these new dynamic optimization tools are illustrated in an optimal hydro-scheduling problem. This chapter contains several novel results, many results have been submitted to publication and can be found in [61]; further material in this chapter will be used for future publications.

Chapter 1

Optimization under Uncertainty

Today nearly every energy management optimization problem suffers from uncertainty to some degree. Once the relevant uncertainties are identified for the optimization model, there are two major classes of approaches to employ: the stochastic programming and the robust optimization methods. In stochastic programming, the uncertain inputs are modeled as discrete or continuous random variables with a known probability distribution. In this thesis we concentrate our study in problems that have continuous random variables. In contrast, the robust optimization approach does not presuppose the existence of a probability distribution for the uncertain parameters; instead the uncertain variables are assumed to take values from some bounded set.

To solve stochastic optimization problems, we transform them into an equivalent deterministic optimization problem and solve them through classical linear or nonlinear optimization techniques. Likewise, in robust optimization, we seek the best robust value of the objective among all robust feasible solutions to the problem, within a deterministic framework. This is why, in the next section, we start with a brief overview of deterministic nonlinear constrained optimization.

1.1 Deterministic constrained optimization

In many applications, there is the need to solve optimization problems not over the whole space $\mathbb{R}$, but rather over a part of it. This part of the space is usually determined explicitly by some linear or non-linear functions, and in such cases one has a constrained optimization problem. For an in-depth study we recommend [95, 10, 112]. A simple prototype optimization problem with deterministic parameters and inequality constraints can be stated as

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_j(x) \leq 0,\ j = 1, 2, \ldots, k. \end{aligned} \tag{1.1}$$

In this formulation, $x \in \mathbb{R}^n$ is the decision vector, the function f is the *objective function* and $g_j : \mathbb{R}^n \to \mathbb{R}$ are the *constraint functions*. Additionally, the set

$$M := \{x \in \mathbb{R}^n : g_j(x) \leq 0 \ \forall j = 1, \ldots, k\} \tag{1.2}$$

is the *feasible set*. A solution to the problem (1.1) is a point $x^* \in M$ such that $f(x^*) \leq f(x)$ for all $x \in M$. It could also be the case that the solution of a given optimization problem is a *set* of solutions, with the understanding that this set could be empty. The set of all minimizers of f on M is denoted $argmin\{f(x) : x \in M\}$.

Definition 1.1. $x^* \in M$ is a *local minimum* of f if there exists $\epsilon > 0$ for which $f(x^*) \leq f(x)$ for any $x \in M \cap B_\epsilon(x^*)$.

Definition 1.2. $x^* \in M$ is a *global minimum* of f over M if $f(x^*) \leq f(x)$ for all $x \in M$.

Local and global maxima are defined similarly. Moreover, it is well-known that x^* is a minimum of $F(x)$ on M if and only if x^* is a maxima of $-F(x)$ on M.
There are three main aims to consider in the general framework of optimization theory. First we ask: when is the existence of a feasible solution guaranteed? The conditions for existence are provided by the following classical Theorem:

Theorem 1.3. ***(Weierstrass Extreme Value Theorem)***
Let $M \neq \emptyset$ be a compact subset of $\mathbb{R}^n$ and $f : M \to \mathbb{R}$ be a continuous function on M. Then there exists x_m and x_M in M such that $f(x_m) \leq f(x) \leq f(x_M)$ for all $x \in M$.

Secondly, we would like to identify conditions that ensure the uniqueness of solution.

Definition 1.4. We say that $f : \mathbb{R}^n \to \mathbb{R}$ is *quasi-convex* if

$$f(\theta x + (1-\theta)y) \leq \max\{f(x), f(y)\} \ \forall x, y \in \mathbb{R}^n,\ \theta \in (0,1).$$

Also $g : \mathbb{R}^n \to \mathbb{R}$ is *quasi-concave* if $-g$ is quasi-convex. And it is *strictly quasi-convex* if

$$f(\theta x + (1-\theta)y) < \max\{f(x), f(y)\} \ \forall x, y \in \mathbb{R}^n,\ \theta \in (0,1).$$

Theorem 1.5. *Suppose M is a non-empty, compact and convex subset of $\mathbb{R}^n$. Suppose $f : M \to \mathbb{R}$ is a continuous and strictly quasi-convex (quasi-concave) function on M. Then, there exists $x^* \in M$ as unique global minimum (maxima).*

One can find a proof of these previous classical theorems, for example, in [112]. Thirdly, we would like to recognize a solution and obtain a characterization of the set of optimal points. This is normally divided in two categories: the identification of conditions that every solution to an optimization problem must satisfy, that is, *necessary conditions* for an optimum point. Then, one identifies conditions such that any point that meets these conditions is a solution, that is, *sufficient conditions* to identify a point as optimal. To develop the necessary and sufficient conditions, let us first introduce the *Lagrangian function*.

Definition 1.6. The Lagrangian function for the problem (1.1) is defined by

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{j=1}^{k} \lambda_j g_i(x),$$

where λ_j (j=1, ..., k) are the Lagrange multipliers.

As we will see throughout the next sections and chapters, the Lagrangian and the Lagrange multipliers are key tools for the development of solution approaches in the framework of constrained problems. Next, we give the Karush-Kuhn-Tucker (KKT) Theorem. Note that this theorem provides necessary conditions that are not sufficient for a minimum. But before doing so, we need to discuss a technical condition called *constraint qualification*, which guarantees that no degenerate behavior occurs at x^*. The most often used one is the linear independence constraint qualification, which is based on the active set.

Definition 1.7. Let x^* be a feasible point of Problem (1.1) and define the active set $E(x^*) := \{i \in \{1, \ldots, k\} : g_i(x^*) = 0\}$, the *linear independence constraint qualification (LICQ)* holds at x^* if the set of active constraint gradients $\{\nabla g_j(x^*)\}_{j \in E(x^*)}$ is linearly independent. That is, if $\sum_{j \in E(x^*)} \lambda_j \nabla g_j(x^*) = 0$ then $\lambda_j = 0$ for all $j \in E(x^*)$. LICQ holds vacuously if $E(x^*) = \emptyset$.

Definition 1.8. [88] Let x^* be a feasible point of Problem (1.1) and let $E(x^*)$ be as in Definition 1.7, the *Mangasarian-Fromovitz constraint qualification (MFCQ)* holds at x^* if there exists a vector $d \in \mathbb{R}^n$ such that

$$\nabla g_j(x^*)^T d < 0 \ \ \forall j \in E(x^*).$$

Theorem 1.9. *[82, 78] Let $f, g_j : \mathbb{R}^n \to \mathbb{R}$ be $C^1(\mathbb{R}^n)$ functions, $j = 1, \ldots, k$. Suppose that x^* is a local solution of* (1.1) *and that a constraint qualification holds at x^*. Then there is a multiplier vector λ^*, with components λ_j^*, $j = 1, \ldots, k$, such that the following conditions are satisfied at (x^*, λ^*)*

i) $\nabla f(x^*) + \sum_{j=1}^{k} \lambda_i^* \nabla g_i(x^*) = 0$,

ii) $g_j(x^*) \leq 0$, *for* $j = 1, \ldots, k$,

iii) $\lambda_j^* g_j(x^*) = 0$ *for* $j = 1, \ldots, k$,

iv) $\lambda_j^* \geq 0$ *for* $j = 1, \ldots, k$.

The conditions $i) - iv)$ *are necessary conditions for a local minimum of problem* (1.1) *and are called the Kuhn-Tucker conditions. For a maximization problem, the non-negativity condition* $iv)$ *is replaced by the condition* $\lambda_j^* \leq 0$.

Remark 1.10. By ii) and iii) we have that if $g_j(x^*) < 0$, then $\lambda_j^* = 0$; that is, for all $i \notin E(x^*)$ we have that $\lambda_i^* = 0$. In other words, the KKT multipliers corresponding to inactive constraints are zero.

Example 1.1. *It should be remarked here that we need the contraint qualification to guarantee that the multiplier* $\lambda \neq 0$. *For instance, let* $f(x) = x$ *and* $g(x) = x^3$. *Then the constraint set is* $M = (-\infty, 0]$ *and the solution is* $x^* = 0$. *However the KKT condition fails since* $\nabla f(x^*) = 1$ *while* $\nabla g(x^*) = 0$, *so there is no* $\lambda \geq 0$ *such that* $\nabla f(x*) = \lambda \nabla g(x^*)$.

In general, if LICQ holds, none of the active constraint gradients can be zero. Moreover, the constraint qualification (CQ), can be stipulated in a variety of ways, all useful but not all equivalent. The interest in having different CQ's comes from having different optimization problems under which certain CQ's are more easily verified than others.

Theorem 1.11. *[47] If* $x^* \in M$ *from* (1.1) *satisfies LICQ, then* x^* *satisfies MFCQ.*

In the following, we present useful cone sets to further understand the structure of the feasible set and the importance of constraint qualifications. For a detailed view on CQ in finite-dimensional spaces we recommend [97].

Definition 1.12. A set $D \subset \mathbb{R}^n$ is called

i) a *cone* if $td \in D$ for all $t \geq 0$ and $d \in D$,

ii) a *convex cone*, if $d_1, d_2 \in D$ implies that $\theta_1 d_1 + \theta_2 d_2 \in D$, $\forall \theta_1, \theta_2 \geq 0$.

Definition 1.13. Let $D \subset \mathbb{R}^n$ be a cone. Then the *polar cone* of D, denoted by D°, is defined as

$$D^\circ := \{p \in \mathbb{R}^n : p^T x \leq 0 \ \ \forall x \in D\}.$$

For example, $(\mathbb{R}^2_+)^\circ = \mathbb{R}^2_-$. Moreover, two types of cones play an important role in the theory of constrained optimization: the normal and tangent cones.

Definition 1.14. Let $D \subset \mathbb{R}^n$ be a nonempty set, and let $\bar{x} \in D$. Then the *normal cone* of D is defined as

$$N_D(\bar{x}) = \{p \in \mathbb{R}^n : p^T(y - \bar{x}) \leq 0 \ \forall y \in D\}.$$

The normal cone is defined by varying y inside D, to find normal vectors p that point out from the set D.

Definition 1.15. Let $D \subset \mathbb{R}^n$ be a nonempty set, and let $\bar{x} \in D$. Then the *tangent cone* of D at $\bar{x}$ is defined as

$$T_D(\bar{x}) = \left\{ p \in \mathbb{R}^n | \exists (x_n)_{n\in\mathbb{N}} \subset D \ (t_k)_{k\in\mathbb{N}} \in (0,\infty) : \lim_{n\to\infty} x_n = \bar{x}, \ \lim_{k\to\infty} \lambda_k(x_k - \bar{x}) = p \right\}.$$

There is a much simpler way to characterize the tangent cone when D is convex as the next proposition states.

Proposition 1.16. *[1, Prop. 2.19] Let $C \subset \mathbb{R}^n$ be a closed nonempty convex set and let $\bar{x} \in C$. Then the tangent and normal cones are closed convex cones and*

$$T_C(\bar{x}) = (N_C(\bar{x}))^\circ = \{p \in \mathbb{R}^n | p^T s \leq 0 \ \forall s \in N_C(\bar{x})\}.$$

Definition 1.17. Let $D \subset \mathbb{R}^n$ be a nonempty set, and let $x \in D$. Then the set of all *feasible directions* of D at x is defined as

$$V_D(x) := \{d \in \mathbb{R}^n : \exists \epsilon > 0, \text{ such that } \ x + td \in D \ \forall t \in [0, \epsilon]\}.$$

Definition 1.18. A direction $d \in \mathbb{R}^n$ is a *descent direction* of the function f at $\bar{x}$ if there exists $\epsilon > 0$ such that

$$f(\bar{x} + td) < f(\bar{x}), \ \ \forall t \in (0, \epsilon].$$

The set of descent directions of f at $\bar{x}$ is denoted by $F(\bar{x})$.

The next result characterizes the descent directions.

Lemma 1.19. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function at a point $\bar{x} \in \mathbb{R}^n$. Then*

i) $\nabla f(\bar{x})^T d \leq 0, \ \forall d \in F(\bar{x})$.

ii) If $d \in \mathbb{R}^n$ satisfies $\nabla f(\bar{x})^T d < 0$, then $d \in F(\bar{x})$.

Having these concepts at hand, suppose further that $M \subset \mathbb{R}^n$ is defined by a finite number of inequality constraints as in (1.2), with $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, k$ given functions, which we assume to be $C^1(\mathbb{R}^n)$. Then the constraint qualifications

can be thought of as conditions imposed on the functions g_i and its derivatives at or around the point x^*, which guarantee that the tangent cone $T_M(x)$ has an explicit representation in terms of the derivatives of the constraint functions. This is important for developing optimality conditions in optimization, since whenever a point x^* is a local solution to the problem

$$\min f(x) \text{ s.t. } x \in M,$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at x^*, it holds that no descent directions around a neighborhood of x^* are feasible, where descent direction d at x^* means $\nabla f(x^*)^T d \leq 0$. Thus if x^* is a local optimum we have

$$\begin{aligned} \nabla f(x^*)^T(x - x^*) \geq & 0, \ \forall x \in M \cap B_\epsilon(x^*) \\ -\nabla f(x^*)(x - x^*) \leq & 0, \ \forall x \in M \cap B_\epsilon(x^*) \\ -\nabla f(x^*) \in & (T_M(x^*))^\circ \\ -\nabla f(x^*) \in & N_M(x^*) \\ 0 \in & \nabla f(x^*) + N_M(x^*). \end{aligned} \tag{1.3}$$

Now, we define the set of first order feasible directions.

Definition 1.20. Given $x^* \in M$, we define the set of *first order feasible directions* for problem (1.1) as

$$\mathcal{F}_M(x^*) := \{d \in \mathbb{R}^n : \nabla g_j(x^*)d \leq 0 \ \forall j \in E(x^*)\}.$$

Lemma 1.21. *[47] For any $x \in M$, $T_M(x) \subset \mathcal{F}_M(x)$.*

Constraint Qualifications allow explicit characterization of the tangent cone. It basically considers the conditions under which the tangent cone is equal to the cone of directions obtained by linearizing all the constraints active at x^*, i.e. when it holds that

$$T_M(x^*) = \mathcal{F}_M(x^*). \tag{1.4}$$

For a detailed characterization of $T_M(x^*)$ for several constraint qualifications we recommend [123, 47].

The Weierstrass Extreme Value Theorem provides sufficient conditions for the existence of optima. Thus, it can be used in conjunction with other necessary conditions, like the previous Theorem 1.9. Otherwise, the first-order conditions tell us how the first derivatives of f and the active constraints g_i are related to each other

at a solution x^*. However, for the direction $w \in \mathbb{R}^n$ for which $\nabla f(x^*)^T w = 0$, we cannot determine from the first derivative whether a move along this direction increases or decreases the value of the objective function f. Thus, we need knowledge of the second derivatives. In order that these conditions be sufficient, the functions involved have to satisfy certain convexity assumption.

Theorem 1.22. *[10, Thm. 11.18] Consider the problem* (1.1) *where f, g_j are $C^2(\mathbb{R}^n)$ functions. Let x^* be a local minimum of the problem and suppose that x^* satisfies the LICQ, then there exist $\lambda_j \geq 0$ such that*

$$\begin{aligned} \nabla_x \mathcal{L}(x^*, \lambda) &= 0 \\ \lambda_j g_j(x^*) &= 0 \text{ for } j = 1, \ldots, k. \\ y^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda) y = y^T &\left[\nabla^2 F(x^*) + \sum_{i=1}^{k} \lambda_i \nabla^2 g_i(x^*) \right] y \geq 0 \end{aligned}$$

for all $y \in \Lambda(x^*)$, where $\Lambda(x^*) := \left\{ d \in \mathbb{R}^n : \nabla g_j(x^*)^T d = 0,\ j \in E(x^*) \right\}$.

Another important particular case of a constraint qualification is that of the convex optimization problems. In the convex framework, every local solution is a global solution and the KKT conditions are both necessary and sufficient, provided a constraint qualification holds. A well-known constraint qualification for the convex case is Slater's constraint qualification.

Definition 1.23. *Slater's constraint qualification* holds for (1.1) if there exists $\hat{x} \in \mathbb{R}^n$ such that $g_j(\hat{x}) < 0$ for all $j \in \{1, \ldots, k\}$.

Theorem 1.24. *If f is convex and the feasible region M is convex, any local solution of* (1.1) *is also a global solution. Moreover, if f and g_j $(j = 1, \ldots, k)$ are differentiable and if Slater's constraint qualification holds, the KKT conditions $i) - iv)$ are necessary and sufficient for x^* and λ^* to define a global solution. If, in addition, f is strictly convex, the global solution is unique.*

For optimality conditions in non-smooth convex programs (see, e.g., [93]).

1.2 Optimization under Uncertainty

Consider (1.1) with uncertain variables $z \in \mathbb{R}^d$. The prototype optimization problem with uncertain parameters is

$$\begin{aligned} \min \quad & f(x, z) \\ \text{s.t.} \quad & g(x, z) \leq 0, \end{aligned} \tag{1.5}$$

where $x \in \mathbb{R}^n$ is the decision vector, $z \in \mathbb{R}^d$ is an uncertain parameter and $g : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^k$ refers to a constraint mapping.

Uncertainty can be defined as the lack of complete knowledge due to, for example, lack of measuring devices to describe a system or variability of outcomes. Thus, uncertainties can be classified in two categories: as measurable and as unknown uncertainties [117]. Measurable uncertainties usually have a well-characterized probability distribution and statistical information like mean, variance, covariance, moments and correlation among input random variables is given. Such a situation arises when historical data for z has been analyzed by staticians or econometrists. In contrast, unknown uncertainties arise from difficulties to measure parameters and there is no given characterization in terms of a probability distribution. In such a case, a model is introduced that ensures feasibility of its solution for all possible realizations (equivalently: for the worst-case scenario) of the uncertain parameter within some uncertainty set.

1.3 Stochastic Programming

Consider the above problem (1.5) and assume that $\xi \in \mathbb{R}^d$ is a continuous random vector with known probability distribution $\mathbb{P}$. Then it becomes an *stochastic programming* problem :

$$\begin{aligned} \min \quad & f(x, \xi) \\ \text{s.t.} \quad & g(x, \xi) \leq 0, \end{aligned} \tag{1.6}$$

where $x \in \mathbb{R}^n$ is the decision vector, $\xi \in \mathbb{R}^d$ is a random vector on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with support on a (closed) set $\mathcal{S} \subset \mathbb{R}^d$, $g : \mathbb{R}^n \times \mathcal{S} \to \mathbb{R}^k$ a constraint mapping and objective function $f : \mathbb{R}^n \times \mathcal{S} \to \mathbb{R}$. There are three classical ways to solve problem (1.6): expected value method, recourse based model and probabilistic constraints. For each of the models, we answer two basic questions:

i) When does it make sense to use the model?

ii) When is it possible to solve analytically or numerically?

The basic idea for solving a stochastic optimization problem is to convert the problem into an equivalent deterministic form, so that the conventional optimization techniques mentioned in Section 1.1 can be applied to it.

1.3.1 Expected value Model

The expected value model is the simplest model of stochastic programming. All the uncertain parameters, whether in the objective function or constraints, are replaced by their expected values. Thus, problem (1.6) is transformed to

$$\begin{aligned} \min \quad & \mathbb{E}\left[f(x,\xi)\right] \\ \text{s.t.} \quad & \mathbb{E}\left[g_j(x,\xi)\right] \leq 0,\ j = 1,2,\ldots,k \end{aligned} \tag{1.7}$$

where $x \in \mathbb{R}^n$ is the decision vector, $\xi \in \mathbb{R}^d$ is a random vector on $(\Omega, \mathcal{F}, \mathbb{P})$ with support on a (closed) set $\mathcal{S} \subset \mathbb{R}^d$. This formulation minimizes the objective function and the constraints on average. Here we assume that $f : \mathbb{R}^n \times \mathcal{S} \to \bar{\mathbb{R}}$ is an integrand such that $f(x,\cdot)$ is measurable and $\mathbb{E}([f(x,\xi)]_{\pm}) < +\infty$. The solution procedure differs and depends on whether the random variable has a continuous distribution or a discrete distribution. In a discrete distribution, there is a finite number of realizations for an uncertain parameter $(\xi_1, \xi_2, \ldots, \xi_m)$ called scenarios and a probability of occurrence $P = (p_1, \ldots, p_m)$ for each scenario. Then, each scenario is like a deterministic model, and we optimize the weighted sum of all scenarios given by:

$$\begin{aligned} \min \quad & \sum_{i=1}^{m} p_i F(x,\xi_i) \\ \text{s.t.} \quad & \sum_{i=1}^{m} p_i g_j(x,\xi_i) \leq 0,\ j = 1,2,\ldots,k. \end{aligned} \tag{1.8}$$

In case of a continuous distribution, which will be the focus of this thesis, the objective and constraint functions are integral functions. To ease notation, we introduce the image (Borel) measure $\mu = \mathbb{P} \circ \xi^{-1}$ on $(\mathbb{R}^d, \mathcal{B}_d)$ and recall the change of variables $\mathbb{E}(\xi(\omega)) = \int_\Omega \xi(\omega)\mathbb{P}(d\omega) = \int_{\mathcal{S}} \xi\ \mu(d\xi)$ where $\mathcal{S} \subset \mathbb{R}^d$. This leads to the following deterministic reformulation of problem (1.7):

$$\begin{aligned} \min \quad & \int_{\mathcal{S}} F(x,z)d\mu(z) \\ \text{s.t.} \quad & \int_{\mathcal{S}} g_j(x,z)d\mu(z) \leq 0,\ j = 1,2,\ldots,k. \end{aligned} \tag{1.9}$$

It makes sense to use this model when the random quantities $g_j(x,z)$ do not have high variability. The optimization of the expected value can be justified by an application of the *Strong Law of Large Numbers*. That is, if we solve problem (1.7), under

the same probability distribution oftentimes this model gives the best possible solution *on average*. If the uncertain parameters have high variability, the obtained solutions through this simple model may not be robust and reliable. Therefore, other stochastic optimization methods should be considered. Moreover, we see that to formulate this model, we need to specify the probability distribution P or μ for the discrete and continuous case. With respect to the second question $ii)$ of solving problem (1.9) numerically, we have to calculate the corresponding Lebesgue integrals. For continuous distributions, this integrals cannot be evaluated numerically with high accuracy for the number of random variables $d > 4$ (see, e.g., [94]).

1.3.2 Recourse model

Recourse models also fall into the category of stochastic programming. They were first introduced in [37, 9]. In a recourse-based model, decisions are taken prior to the realization of the random variables, then recourse or corrective measures are taken to compensate for possible violation of constraints. We proceed with the presentation of the two-stage and multi-stage stochastic programs with recourse. For a more general and exhaustive introduction to recourse stochastic programming, we refer the reader to [19, 107, 122].

Two-stage stochastic programming

The most basic stochastic recourse problem is the *two-stage stochastic program with recourse*. To state the problem, assume that *non-anticipative decisions* represent the main decisions that have already been made and that a temporary violation of the random constraint is allowed. Feasibility is restored through recourse actions that are deferred until the realization of uncertainty is observed. In this way, the decisions are partitioned into two-stages:

1. *First stage* or *here-and-now* decisions. These decisions are made before the realization of the random vector occurs. Thus, variables representing here-and-now decisions do not depend on each realization of the stochastic parameter.

2. *Second stage* or *wait-and-see* decisions. These decisions are made after knowing the actual realization of the random vector $\xi(\omega)$. Therefore, these decisions depend on each realization of the random vector.

We denote the first-stage decisions by the vector $x \in \mathbb{R}^{n_1}$, and the second-stage decisions by the vector $y(x,\xi) \in \mathbb{R}^{n_2}$. Then problem (1.6) takes the following deterministic form

$$\begin{aligned} \min_{x} \quad & f_1(x) + \mathcal{Q}(x) \\ \text{s.t.} \quad & g_{1,j}(x) \leq 0,\ j = 1,2,\ldots,k \end{aligned} \tag{1.10}$$

where $\mathcal{Q}(x) = \mathbb{E}[Q(x,\xi)]$, and after the realization of $\xi(\omega)$ becomes available, an optimal decision about $y(x,\xi(\omega))$ is made by solving the *recourse problem:*

$$\begin{aligned} Q(x,\xi) = \min_{y} \quad & f_2(x,y,\xi) \\ \text{s.t.} \quad & t_{2,j}(x,\xi) + g_{2,j}(y(\xi),\xi) \leq 0,\ j = 1,\ldots,k. \end{aligned} \tag{1.11}$$

where $f_2 : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathcal{S} \to \mathbb{R}$, $t_{2,j} : \mathbb{R}^{n_1} \times \mathcal{S} \to \mathbb{R}$, $g_{2,j} : \mathbb{R}^{n_2} \times \mathcal{S} \to \mathbb{R}$ for all $j = 1,\ldots,k$. We assume that all functions $f_2(\cdot,\xi), t_{2,i}(\cdot,\xi)$ and $g_{2,i}(\cdot,\xi)$ are continuous for any fixed ξ, and measurable in ξ for any fixed argument. Given this assumption, $Q(\cdot,\xi)$ is measurable. One takes first-stage decisions that are on average optimal, with the possibility to take recourse decisions afterwards to counter the effects of the revealed uncertainty. The difficulty inherent in this approach lies in the computational burden of computing $\mathcal{Q}(x)$ for all x.

Feasible sets

In addition, it is convenient to define two feasible sets associated with different stages of the recourse model. First, let M_1 be the set that does not depend on the particular realization of the random vector:

$$M_1 := \left\{x : g_{1,j}(x) \leq 0, j = 1,\ldots,k\right\}.$$

Similarly, the second-stage feasible set is given by

$$M_2 := \left\{x : \mathcal{Q}(x) < \infty\right\}.$$

Thus, the two-stage stochastic programming problem can be reformulated as

$$\begin{aligned} \inf \quad & F_1(x) + \mathcal{Q}(x) \\ \text{s.t.} \quad & x \in M_1 \cap M_2. \end{aligned} \tag{1.12}$$

We also introduce the elementary feasible sets, associated with the realizations of the random vector, defined as

$$M_2(\xi) := \left\{ x : \exists y(\xi) \ s.t. \ t_{2,j}(x,\xi) + g_{2,j}(y(\xi),\xi) \leq 0, \ j = 1,\dots,k. \right\}.$$

Convexity

In the following we present some of the conditions that guarantee that the two-stage stochastic program is convex and well-defined. For more details (see, e.g., section 3.4 [19]).

Proposition 1.25. *If $f_1, g_{1,j}$ $(j = 1,\dots,k)$ are convex on $\mathbb{R}^{n_1}$, $f_2(\cdot,\xi), g_{2,j}(\cdot,\xi)$ $(j = 1,\dots,k)$ are convex on $\mathbb{R}^{n_2}$ for all $\xi \in \mathcal{S}$ and $t_{2,j}(\cdot,\xi)$ $(j = 1,\dots,k)$ are convex on $\mathbb{R}^{n_1}$ for all $\xi \in \mathcal{S}$ then* (1.12) *is a convex optimization problem.*

It is assumed that (1.11) satisfies the following Slater's constraint qualification (see Definition 1.23) for the second-stage problem.

$$\begin{gathered} \textit{If } \mathcal{Q}(x) < \infty \textit{ for almost all } \xi \in \mathcal{S}, \textit{ there exists some } y(\xi) \textit{ such that} \\ t_{2,j}(x,\xi) + g_{2,j}(y(\xi),\xi) < 0 \textit{ for } j = 1,\dots,k. \end{gathered}$$

This ensures that the second-stage problem is well-defined and that a solution can be obtained by solving the associated KKT system. We also obtain continuity of the recourse function if we assume that M_2 is bounded, as stated in the following theorem. The proof can be found in [19], p. 124.

Theorem 1.26. *If the recourse feasible region is bounded for any $x \in \mathbb{R}^{n_1}$, then the function $Q(x,\xi)$ is lower semi-continuous in x for all $\xi \in \mathcal{S}$.*

The following result ensures the existence of a solution. The proof can be found in [19], page 126.

Theorem 1.27. *Assume that the two-stage stochastic program is convex, and that f_1 is continuous, $g_{1,j}, g_{2,j}$ are continuous for each $j = 1,\dots,k$. Suppose further that the recourse feasible region is bounded for any $x \in \mathbb{R}^{n_1}$, M_1 is bounded, $M_1 \cap M_2 \neq \emptyset$. Then* (1.10)-(1.11) *have a finite optimal solution and the infimum of $f_1(x) + \mathcal{Q}(x)$ is attained.*

For optimality conditions in non-smooth convex stochastic programs with recourse (see, e.g. [16]).

Multi-stage stochastic programming

The *multi-stage stochastic program with recourse* relies intuitively on the same ideas as the two-stage version. Decisions are made without anticipating future realizations of uncertainty and are therefore partitioned into stages according to the information flow. The realization of uncertain data is gradually revealed and decisions are taken *dynamically*. However, a crucial difference in multistage decision making is the available information, which is represented in an increasing sequence of σ-algebras: a filtration. Many problems in energy management involve a sequence of decisions that respond to realizations of uncertain outcomes that are not known a priori. In particular, this is useful for optimal power dispatch problems (see, e.g., [64, 116]).

In this framework, we formulate the problem by introducing measurability conditions to state decisions that, at a certain stage, depend only on the available information at that point in time. Uncertainty is described by a discrete-time stochastic process $\xi = (\xi_1, \xi_2, \dots, \xi_T)$ and specifically revealed uncertainty available at time t is denoted by $\xi_{[t]} = (\xi_1, \dots, \xi_t)$. We denote by $\mathcal{F}_t \subseteq \mathcal{F}$ the $\sigma-$algebra generated by $\xi_{[t]}$ and assume that the $\sigma-$algebras form a filtration such that $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$ for $t = 1, \dots, T-1$ and $\mathcal{F}_1 = \{\emptyset, \Omega\}$ and $\mathcal{F}_T = \mathcal{F}$. *Non-anticipativity constraint* means that t-th stage decision x_t is $\mathcal{F}_{t-1}$-measurable, i.e., $x_t = \mathbb{E}(x_t|\mathcal{F}_{t-1})$. At stage t, we know $\xi_{t-1}(\omega)$ as well as x_{t-1}. We have to decide on x_t such that the constraints are satisfied. We assume that in the first step, the objective function is deterministic, so the multi-stage program can be formulated as the following nested optimization problem

$$\begin{aligned}
\min_{x_1} \quad & f_1(x_1) + \mathbb{E}_{\xi_1}\left(\min_{x_2} f_2(\xi_{[1]}, x_{[1]}, x_2(\xi_1))\right) + \mathbb{E}_{\xi_2|\xi_1}\left(\min_{x_3} f_3(\xi_{[2]}, x_{[2]}, x_3(\xi_2))\right) + \cdots \\
& + \mathbb{E}_{\xi_T|\xi_1,\dots,\xi_{T-1}}\left(\min_{x_T} f_T(\xi_{[T-1]}, x_{[T-1]}, x_T(\xi_{T-1}))\right) \\
\text{s.t.} \quad & g_{1,j}(x_1) \leq 0, \quad j = 1, 2, \dots, k., \\
& g_{t,j}(\xi_{[t]}, x_{[t-1]}(\xi_{[t-2]}), x_t(\xi_t)) \leq 0, \quad j = 1, 2, \dots, k.
\end{aligned} \tag{1.13}$$

The increase of the number of stages raises the difficulty of finding an optimal solution.

It makes sense to use the recourse model to solve a dynamic optimization problem characterized by uncertain future outcomes of some input data. The objective of the model is to come up with fist-stage decisions that are robust enough such that uncertain events can be tackled by taking different recourse actions; i.e., the optimal solution hedges against all possible events that may arise in the future. The sequence

of events and decisions is thus summarized as:

$$x_1 \curvearrowright \xi_1 \curvearrowright x_2(\xi_1, x_1) \curvearrowright \cdots \curvearrowright \xi_{t-1} \curvearrowright x_t(\xi_{[t-1]}, x_{[t-1]}) \tag{1.14}$$

where we take first stage decision x_1, after the random variable is revealed, we take second-stage decision x_2, then the random variable is revealed, and we end up with a decision x_t. The latter is denoted by $x_t(\xi_{[t-1]}, x_{[t-1]})$, to stress that the last decision x_t differs according to the outcomes of the random experiment $\xi_{[t-1]}$ and of the previously made decisions $x_{[t-1]}$.

Due to the computational difficulties of calculating the expectations in problem (1.13), which involve multi-dimensional integration of implicitly defined functions, a standard approach to solving the recourse problem is by constructing scenarios. That is, the probability distribution is mostly approximated by a discrete distribution with finite support. We assume that the approximate distribution of the stochastic process $\xi = (\xi_i)_{i=1}^T$ is given by scenario paths $\xi^s = (\xi_i^s)_{i=1}^T$, $s = 1, \ldots, S$, and the scenario probabilities p_s, $s = 1, \ldots, S$. This concept allows one to represent the possible outcomes by means of a graph $\mathcal{G} = \{v_1, \ldots, v_T\}$ called *scenario tree* with node v. Thus, the branching occurs with the arrival of new information at stage $t \in \{1, \ldots, T\}$. Let $\mathcal{D}(v)$ be the set of decendants of $v \in \mathcal{G}$ and $\mathcal{A}(v)$ the set of its ancestors. The root of the tree is denoted by r and $d(v)$ represents the depth of $v \in \mathcal{G}$, i.e. $d(r) = 0$ and $d(v_t) = t$. Each node v is associated to a conditional probability $\pi(v)$ to be chosen at the stage $d(v) - 1$ and $\pi(r) = 1$. Assume that the random variables are independent; then the probability of scenario s is

$$P^s = \prod_{i=0}^{T} \pi(v_i^s).$$

The probability of the path leading to a node v is

$$\mu_v = \pi(v) \prod_{w \in \mathcal{A}(v)} \pi(w).$$

Let x_v be the decision vector associated to node $v \in \mathcal{G}$, and $x_{[v]} = \{x(w) : w \in \mathcal{A}(v) \cup \{v\}\}$ be the decisions made at previous stages. If the objective and constraints are separable between the different stages, the multistage stochastic program

with recourse can be expressed as

$$\begin{aligned} \min \quad & \sum_{v \in \mathcal{G}} \mu_v f_v(\xi_{[d(v-1)]}, x_{[v-1]}, x_v) \\ \text{s.t.} \quad & g_j(\xi_{[d(v-1)]}, x_{[v-1]}, x_v) \leq 0, \quad j = 1, \ldots, k. \end{aligned}$$

1.3.3 Optimization under Probabilistic Constraints

The third type of stochastic optimization model is called *Probabilistic Constraint Programming* (PCP). This model belongs to one of the major approaches to dealing with uncertain parameters in optimization problems, and it is the main focus of this thesis. In contrast to the recourse problem, it is of static type, where the decision variables are independent of time. Moreover, the recourse-based approach to stochastic programming allows infeasibilities in the second stage at a certain cost. The approach focuses on the minimization of expected recourse costs. In the probabilistic constraint approach, the focus is on the system's capability to meet feasibility in an uncertain environment. The prototype of *probabilistic constraint programming* has the form

$$\begin{aligned} \min \quad & C(x) \\ \text{s.t.} \quad & \mathbb{P}\left(g\left(x, \xi\right) \geq 0\right) \geq p \\ & x \in \mathcal{X}. \end{aligned} \tag{1.15}$$

Here $\mathcal{X} \subset \mathbb{R}^n$ is a nonempty set, $C : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^k$ a constraint mapping, and ξ is a d-dimensional random vector defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The value $p \in (0, 1)$ is called the probability level, and it is specified by the decision maker in order to model the safety, risk or reliability of the decision x. The meaning of the *probabilistic constraint* is as follows: a decision x is declared to be feasible, whenever the probability of satisfying the k-dimensional random inequality system is at least p.

One can use the so-called probabilistic or chance constrained optimization model when the restrictions are affected by random parameters, and when a decision maker has to provide an optimal decision before observing the realization of these random parameters. Some of the important applications of probabilistic constrained programming include engineering [86], water resource management [43, 2], resource management [17], telecommunications [34], insurance [23, 80] and energy economics [145].

A probabilistic constrained model was first introduced in [32] and [31] as a tool to solve linear programming model for determining refinery rates for heating oils

to meet stochastic demands. Since then, many contributions have been made to the structural analysis and numerical methods. Among these are the important structural and stability contributions from Prékopa ([103, 104, 106, 108]). Other key contributions include [136], [69, 72, 73], [74, 115] and [41]. For basic monographs on this topic, we refer to [107] and [122, ch. 4].

Conceptual Framework

There exist two conceptually different models for optimization problems under probabilistic constraints. Thus, there are two ways in which we can transform problem (1.6).

The type of probabilistic constraint shown in (1.15) can be explicitly written as

$$\mathbb{P}\left(g_j\left(x,\xi\right) \geq 0 \ \forall j \in \{1,\ldots,k\}\right) \geq p \tag{1.16}$$

and it is referred to as *Joint Probabilistic Constraint* (JPC). This type of probabilistic constraints was first formulated in [91]. For the special case of a random right-hand-side vector whose elements are independent random variables, they show that a deterministic equivalent of a JPC model is a concave programming problem.

Alternatively, when [32] and [31] originally introduced probabilistic constrained programming, they dealt with *Individual Probabilistic Constraints*. Under this concept, one turns each component of the stochastic inequality system into several single probabilistic constraints, thereby having a set of individual probabilistic constraints of the form

$$\mathbb{P}\left(g_j(x,\xi) \geq 0\right) \geq p_j \ \forall j \in \{1,\ldots,k\} \tag{1.17}$$

with possibly different and arbitrary probabilities levels $p_j \geq p$. Such *individual probabilistic constraints* though yielding a larger system of k inequalities as compared to one inequality in the joint case, lead to a loss of robustness, which is compensated by a much simpler numerical and theoretical treatment. If one is looking for decisions such that the restrictions of the whole stochastic inequality system are satisfied at the given probability level, then a formal solution via the individual model, may result in completely unreliable optimal decisions. In [144] the authors consider a problem from water reservoir management; they show how the use of simple individual probabilistic constraints can completely fail, while the use of joint probabilistic constraints yields robust solutions. In this chapter, we provide an overview of the known results on the structural properties of joint probabilistic constraints of the form (1.16). It will lay a foundation for the theoretical extensions of *dynamic*

joint probabilistic constraints, which are presented in Chapter 4. In this framework, problem (1.6) can be reformulated as

$$\begin{aligned} \min \quad & C(x) \\ \text{s.t.} \quad & \mathbb{P}\left(g_j\left(x,\xi\right) \geq 0 \ \forall j \in \{1,\ldots,k\}\right) \geq p. \end{aligned} \tag{1.18}$$

Remark 1.28. Traditionally, under PCP, the objective function in problem (1.6) is replaced by the expected value $C(x) := \mathbb{E}(f(x,\xi))$. This approach is risk-neutral and will be the focus of our applications in this thesis. However, one can also model a risk averse objective function using risk measures (see, e.g. , ch. 6 [122]) or [7]).

Measurability

We start by discussing the well-foundedness of a probability constraint of the type (1.15). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For a fixed $x \in \mathbb{R}^n$, is the set

$$\left\{\omega \in \Omega : g_j\left(x,\xi(\omega)\right) \geq 0 \ \forall j \in \{1,\ldots,k\}\right\}$$

measurable? Such that we can JPC talk about probabilistic functions

$$\mathbb{P}\left(\left\{\omega \in \Omega : g_j\left(x,\xi(\omega)\right) \geq 0 \ \forall j \in \{1,\ldots,k\}\right\}\right)?$$

This is positively answered through the following two lemmas, which we prove.

Lemma 1.29. *Let $h : \mathbb{R}^n \to \mathbb{R}$. The level set $L_t := \{x \in \mathbb{R}^n | h(x) \geq t\}$ is closed for all $t \in \mathbb{R}$ if and only if h is an upper semi-continuous function.*

Proof. Assume that $t \in \mathbb{R}$ and $\{x_n\} \subset L_t$ is any sequence such that $x_n \to x^*$, with $x^* \in \mathbb{R}^n$. Since h is upper semi-continuous it follows

$$h(x^*) \geq \limsup_{x \to x^*} h(x) \geq \limsup_{n \to \infty} h(x_n) \geq t,$$

so $x^* \in L_t$ and L_t is therefore closed. Conversely, let $x^* \in \mathbb{R}^n$ be arbitrary and pick a sequence $x_n \to x^*$. Since L_t is a closed set, we have that $L_t^c = \{x \in \mathbb{R}^n | h(x) < t\}$ is open. So let $t > h(x^*)$. By definition, L_t^c contains an open neighborhood around x^* with radius $\epsilon > 0$, which we denote $B_\epsilon(x^*)$. Hence

$$\limsup_{x_n \to x^*} h(x_n) = \lim_{\epsilon \to 0} \sup_{x \in B_\epsilon(x^*)} h(x_n) \leq t.$$

Since this is true for every $t > h(x^*)$, we have that

$$\limsup_{x_n \to x^*} h(x_n) \leq h(x^*).$$

Therefore, h is upper semi-continuous. □

Lemma 1.30. *Let $\xi \in \mathcal{R}^d$ be a random variable. Consider a function $g : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^k$ such that for any $x \in \mathbb{R}^n$ each coordinate $g_i(x, \cdot)$ is upper-semicontinuous. Then for any $x \in \mathbb{R}^n$ the set $\{\omega \in \Omega : g(x, \xi(\omega)) \geq 0\}$ is Borel measurable.*

Proof. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space such that $\xi : (\Omega, \mathcal{F}) \to (\mathbb{R}^d, \mathcal{B}^d)$ is a measurable function. Let us denote $g_x := g(x, \cdot)$ and $g_{i,x} := g_i(x, \cdot)$. We have that

$$\{\omega \in \Omega : g(x, \xi(\omega)) \geq 0\} = (g_x \circ \xi)^{-1} ([0, \infty)).$$

We claim that $(g_x \circ \xi)^{-1} ([0, \infty)) = \xi^{-1}(g_x^{-1}[0, \infty))$ is measurable. Indeed, as ξ is a random variable, we only have to prove that $g_{i,x}^{-1}([0, \infty))$ is Borel measurable for all $i \in \{1, \ldots, k\}$. Since $g_{i,x}$ is upper semi-continuous for all $i \in \{1, \ldots, k\}$, by Lemma 1.29 the level sets L_t are closed for all t. In particular, $L_0 = g^{-1}([0, \infty))$ is a closed set, thus it is Borel measurable. □

Measurability cannot be so easily assured. It is not sufficient, for instance, that g is concave, since a convex set need not be Borel measurable. To see this, one can construct convex sets in $\mathbb{R}^n$ such that the boundary is non-Borel measurable (for existence of such a set, see, for example, [77, p. 125]). On the other hand, a convex set is Lebesgue-measurable, since the boundary has Lebesgue measure zero. Thus, if ξ admits a density with respect to the Lebesgue measure and g is concave or upper semi-continuous in the second argument, then (1.16) is well-founded.

Closedness of the Feasible Set

Compactness of M_p is of interest for theoretical and algorithmic purposes. First, we know by Weierstrass Theorem that nonempty and compact feasible sets guarantee the existence of solutions for an optimization problem in $\mathbb{R}^n$. Secondly, compact feasible sets allow us to derive stability results for solutions when the usually unknown distribution of ξ has to be approximated on the basis of estimations and historical observations [73].

In this section, we ask two fundamental questions. First, under what conditions is the feasible set

$$M_p := \{x \in \mathbb{R}^n : \mathbb{P}(g(x, \xi) \geq 0) \geq p\}$$

closed? This property is shown in [70], but here we present a different proof.

Lemma 1.31. *Let $\xi \in \mathcal{B}^d$ be a random variable. Assume further that $g_j : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ is upper semi-continuous in both variables for all $j \in \{1, \dots, k\}$. Then, the mapping $\varphi : \mathbb{R}^n \to [0,1]$,*

$$\varphi(x) := \mathbb{P}\left(g_j(x,\xi) \geq 0 \ \forall j \in \{1,\dots,k\}\right) \tag{1.19}$$

is upper semi-continuous. As a consequence, the set $M_p := \{x \in \mathbb{R}^n : \varphi(x) \geq p\}$ is closed for all $p \in [0,1]$.

Proof. Let $x^* \in \mathbb{R}^n$ and pick a sequence $\{x_n\} \subset \mathbb{R}^n$ such that $\lim_{n\to\infty} x_n = x^*$ a.s. Consider the sets

$$A_n := \{\omega \in \Omega : g_i(x_n, \xi(\omega)) \geq 0 \ \forall i = 1, \dots, k.\}$$

$$A := \{\omega \in \Omega : g_i(x^*, \xi(\omega)) \geq 0 \ \forall i = 1, \dots, k.\}.$$

Then, by Fatou's Lemma (see e.g. p.144 [46])

$$\begin{aligned} \mathbb{E}(\liminf \mathbb{1}_{A_n}) &\leq \liminf \mathbb{E}(\mathbb{1}_{A_n}) = \liminf \mathbb{P}(A_n) \\ &= \liminf \varphi(x_n) \leq \limsup \varphi(x_n) \\ &= \limsup \mathbb{P}(A_n) = \limsup \mathbb{E}(\mathbb{1}_{A_n}) \\ &\leq \mathbb{E}(\limsup \mathbb{1}_{A_n}). \end{aligned} \tag{1.20}$$

Now, let us show that

$$\limsup \mathbb{1}_{A_n} \leq \mathbb{1}_A. \tag{1.21}$$

If we pick $\omega \in A$ then $\mathbb{1}_A = 1$ so (1.21) follows trivially. Pick $\omega \notin A$, this means $g_i(x^*, \xi(\omega)) < 0$ for all $i = 1, \dots, k$. Thus, by the upper-semicontinuity of g_i in the first variable and the convergence almost surely, we have that

$$\limsup_{n\to\infty} g_i(x_n, \xi(\omega)) \leq g_i(x^*, \xi(\omega)) < 0 \ \forall i = 1, \dots, k.$$

So, for large enough N the following inequality is satisfied

$$g_i(x_N, \xi(\omega)) < 0 \ \ \forall i = 1, \dots, k.$$

This implies that $\limsup \mathbb{1}_{A_N}(\omega) = 0$ and by (1.20) we have that

$$\varphi(x_n) \geq \limsup \varphi(x^*)$$

Finally, by similar arguments and using the upper semi-continuity of g, we have that $\mathbb{1}_{A'} \leq \liminf \mathbb{1}_{A_n}$, where

$$A^{'} := \{\omega \in \Omega : g_i(x_n, \xi(\omega)) < 0 \ \forall i = 1, \ldots, k.\},$$

and $\mathbb{1}_{A'} = \mathbb{1}_A$ a.s., which by (1.20) implies the first result. It follows from Lemma 1.29 that M_p is closed for $p \in [0,1]$. □

Secondly, we want to explore when the feasible set

$$M_p = \left\{x \in \mathbb{R}^n : \mathbb{P}\left(g_j(x,\xi) \geq 0 \ \forall j \in \{1,\ldots,k\}\right) \geq p\right\} \subset \mathbb{R}^n$$

is bounded? In general, JPC optimization problems are not bounded, as our next example shows.

Example 1.2. *Let $\xi \sim \mathcal{N}(0,1)$ and consider the static probabilistic constraint problem*

$$\begin{aligned} &min \quad -x \\ &s.t. \quad \varphi(x) = \mathbb{P}\left(\xi \leq \frac{1}{x+1}\right) \geq 0.4. \end{aligned} \tag{1.22}$$

Since $\lim_{x\to\infty} \varphi(x) = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 0.5$, for large values of x the constraint is satisfied, so the problem is unbounded.

Continuity of Probabilistic Constraints

To find the numerical solution of an optimization problem of the type 1.18 it is crucial to have structural properties for the Joint probabilistic constraint φ. For instance, continuity is a necessary condition for the direct search methods (see, e.g. [101, 149]). In the previous section, we concluded that from the upper semi-continuity of g, we can get the upper semi-continuity of φ defined in (1.19). However, in the following example, which can be found in [70], we see that lower semi-continuity or continuity cannot be transferred on this way.

Example 1.3. *Let $\xi \sim \mathcal{N}(0,1)$ and consider the affine linear restriction function $g : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2$ given by $g_1(x,\xi) = 2x_1 + x_2 - \xi$ and $g_2(x,\xi) = -x_1 + x_2 + 0.5$.*

$$\varphi(x) = \mathbb{P}\left(g_i(x,\xi) \geq 0 \ i = 1,2\right) = \mathbb{P}\left(\xi \leq 2x_1 + x_2, -x_1 + x_2 \geq -0.5\right).$$

Although g_1 and g_2 are continuous functions and ξ has a continuous probability distribution, the probabilistic constraint $\varphi(x)$ is discontinuous. (see Figure 1.1).

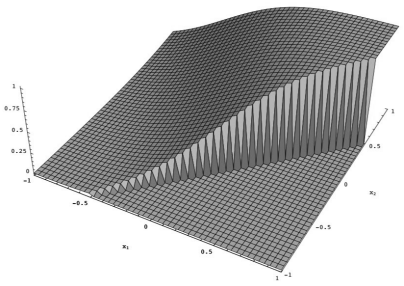


FIGURE 1.1: Example of a discontinuous probabilistic constraint with continuous input data illustrated in [70] .

To show continuity of the probabilistic constraint we must require that no probability mass is given to the boundary of the set $\{z \in \mathbb{R}^d : g(x,z) = 0\}$.

Lemma 1.32. *[70, 111]. Let the constraint mapping $g : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^k$ be jointly lower semi-continuous. Assume further that*

$$\mathbb{P}\left(g_i(x,\xi) = 0\right) = 0 \;\; \forall x \in \mathbb{R}^n \; \forall i \in \{1,\ldots,k\}.$$

Then the mapping $\varphi : \mathbb{R}^n \to [0,1]$ defined as $\varphi(x) = \mathbb{P}\left(g_i(x,\xi) \geq 0 \, \forall i = 1,\ldots,k\right)$ is also continuous.

Remark 1.33. If $\xi \in \mathbb{R}^d$ admits a density $f_\xi : \mathbb{R}^d \to \mathbb{R}^+$ with respect to the Lebesgue measure λ, then we can define the probability function φ as the Lebesgue integral

$$\varphi(x) = \int_{\{\omega \in \Omega : g(x,\xi(\omega)) \geq 0\}} f_\xi(u)\lambda(du). \tag{1.23}$$

Theorem 1.34. *Let the constraint mapping $g : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^k$ be jointly continuous. Suppose that random variable $\xi \in \mathbb{R}^d$ admits a density with respect to the Lebesgue measure. Assume further that for all $x \in \mathbb{R}^n$ and for all $i \in \{1,\ldots,m\}$ the sets*

$$\left\{z \in \mathbb{R}^d : g_i(x,\xi) = 0\right\}$$

are Lebesgue null-sets. Then the mapping $\varphi : \mathbb{R}^n \to [0,1]$ defined in (1.23) *is also continuous.*

Proof. Since ξ admits a density f_ξ, it follows that

$$\mathbb{P}(g_i(x,\xi) = 0) = \int_{\{g_i(x,z)=0\}} f_\xi(z)dz = 0 \;\; \forall x \in \mathbb{R}^n \;\; \forall i \in \{1,\ldots,k\}.$$

We use Lemma 1.29 and Lemma 1.32 to conclude the proof. □

Convexity

In addition to analytical properties such as continuity and differentiability of constraint functions, the convexity properties of the set of feasible decisions defined by it $M_p = \{x \in \mathbb{R} : \varphi(x) \geq p\}$ is fundamental for an optimization problem.

In probabilistic constrained programming, the first question one deals with is the convexity of the feasible set defined by the simplest right-hand side probabilistic constraint $g(x,z) = x - z$, in which the constraint in 1.18 directly relates to the distribution function of ξ

$$M_p = \{x \in \mathbb{R} : \mathbb{P}(\xi \leq x) \geq p\} = \{x \in \mathbb{R} : F_\xi(x) \geq p\}.$$

We know when M_p is convex, F_ξ is a quasi-concave function. Indeed, one dimensional distribution functions can never be concave or convex (due to being bounded by zero and one) it turns out that many of them are quasi-concave. More generally, as our next lemma shows, it is sufficient that the function $\varphi(x) = \mathbb{P}(g(x,\xi) \geq 0)$ is quasi-concave for the feasible set M_p to be convex.

Lemma 1.35. *Consider problem 1.15, define $\varphi : \mathbb{R}^n \to [0,1]$ as $\varphi(x) = \mathbb{P}(g(x,\xi) \geq 0)$ and let $M_p = \{x \in \mathbb{R} : \mathbb{P}(g(x,\xi) \geq 0) \geq p\}$. The feasible set M_p is convex for all $p \in [0,1]$ if and only if φ is quasi-concave.*

Proof. Let $p \in [0,1]$, $x, y \in M_p$, and $\alpha \in [0,1]$. Let $x^\alpha = \alpha x + (1-\alpha)y$. Quasi-concavity of φ implies $\varphi(x^\alpha) \geq \min\{\varphi(x), \varphi(y)\} \geq p$, showing $x^\alpha \in M_p$. For the other implication, let $x, y \in \mathbb{R}^n$ and $\alpha \in [0,1]$. Define $p := \min\{\varphi(x), \varphi(y)\}$. It follows that $x, y \in M_p$. Thus, from convexity of M_p we obtain $x^\alpha \in M_p$, that is, $\varphi(x^\alpha) \geq p$. □

Quasi-concavity arises in many distribution functions. Prékopa introduced the notion of log-concave measures and proves the following renowned Theorem 10.2.1 in [107].

Theorem 1.36. *If the functions $g_j(x,z)$ in problem 1.18 are quasi-concave in $\mathbb{R}^{n+d}$ for all $j = 1,\ldots,k$ and the random vector ξ has a continuous distribution with a log-concave density (i.e. a density whose logarithm is a possibly extended-valued concave function). Then $\varphi(x)$ is a log-concave function in $\mathbb{R}^n$. As a consequence, $M_p = \{x \in \mathbb{R} : \mathbb{P}(g(x,\xi) \geq 0) \geq p\}$ is convex for any $p \in [0,1]$.*

The power of this result becomes evident in combination with the following theorem:

Theorem 1.37. *[105, Theorem 2] Let $\mathbb{P}$ be a probability measure in $\mathbb{R}^n$ generated by a log-concave probability density function, $f(x) = e^{-Q(x)}$, $x \in \mathbb{R}^n$ and Q convex function. Then $\mathbb{P}$ is a log-concave probability measure.*

This result is quite strong since many common densities are log-concave; in particular all Gaussian densities. For instance, any ξ with log-concave density together with the linear restriction $g(x, \xi) = Ax - B\xi$, where concavity of the g_j holds true make up a convex feasible set. Nevertheless, for the constraint map where random vectors are separated from decision vectors, like $g(x, \xi) = h(x) - \xi$ (where $g : \mathbb{R}^n \to \mathbb{R}^k$), even if ξ has a log-concave law, we cannot assure quasi-concavity of $h(x) - \xi$. Unlike concavity, quasi-concavity is not preserved under addition, so quasi-concavity of the components h_j would not be enough to ensure convexity of the corresponding feasible set (see, e.g., [75, Example 1.1]).

To derive convexity results for the feasible set when relaxing the strong requirement of concavity in both variables, there is a generalized convexity concept, i.e. α-concavity, which parametrizes concavity properties between concavity and quasi-concavity. The following generalization of α-concavity was suggested by [24].

Definition 1.38. A non-negative function $f(x)$ defined on a convex set $D \subset \mathbb{R}^n$ is said to be **α-concave**, where $\alpha \in [-\infty, \infty]$, if for all $x, y \in D$ and all $\lambda \in [0, 1]$ the following holds: if $\alpha = -\infty$, then

$$f(\lambda x + (1 - \lambda)y) \geq \min\{f(x), f(y)\};$$

if $\alpha = 0$, then

$$f(\lambda x + (1 - \lambda)y) \geq f^{\lambda}(x) f^{1-\lambda}(y);$$

if $\alpha = \infty$ then

$$f(\lambda x + (1 - \lambda)y) \geq \max\{f(x), f(y)\};$$

For any other value of α

$$f(\lambda x + (1 - \lambda)y) \geq [\lambda f^{\alpha}(x) + (1 - \lambda) f^{\alpha}(y)]^{1/\alpha}.$$

In the case $\alpha = 0$, the function is called *log-concave*, for $\alpha = 1$ it is *concave*, and $\alpha = -\infty$ it is *quasi-concave*.

Remark 1.39. If f is α-concave, then it is β-concave for all $\beta \leq \alpha$. Thus, all α-concave functions are quasi-concave.

Definition 1.40. A probability measure $\mathbb{P}$ defined on the Borel subsets of a convex set $\mathcal{C} \subset \mathbb{R}^d$ is said to be α-concave if for all Borel measurable subsets $A, B \subset \mathcal{C}$ and

$\theta \in [0,1]$ we have

$$\mathbb{P}\left(\theta A + (1-\theta)B\right) \geq \left(\theta[\mathbb{P}(A)]^{\alpha} + (1-\theta)[\mathbb{P}(B)]^{\alpha}\right)^{1/\alpha} \tag{1.24}$$

where $\theta A + (1-\theta)B := \{\theta x + (1-\theta)y : x \in A, y \in B\}$ is the Minkowski sum of A and B.

It is clear that if a random variable ξ induces an α-concave probability measure on $\mathbb{R}^d$, then its distribution function $F_{\xi}(x) = \mathbb{P}(\xi \leq x)$ is an α-concave function. Moreover, all non-degenerate quasi-concave probability measure must have a density as the following result shows.

Theorem 1.41. *[22] If $\mathbb{P}$ is a quasi-concave measure on $\mathbb{R}^d$ and the dimension of its support is d, then $\mathbb{P}$ has a density with respect to the Lebesgue measure.*

The following is the most general concavity result for probabilistic constraints. It is proven in [122, ch. 4] or in [22].

Theorem 1.42. *Let $g : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^k$ be a (jointly) quasi-concave function and let $\xi \in \mathbb{R}^d$ be a random variable inducing an α-concave probability distribution $\mathbb{P}$. Then the mapping $x \in \mathbb{R}^n \mapsto \varphi(x) = \mathbb{P}(g(x,\xi) \geq 0)$ is an α-concave function on the set $M = \{x \in \mathbb{R}^n : \exists z \in \mathbb{R}^d \ s.t. g(x,z) \geq 0\}$.*

Remark 1.43. Theorem 1.36 and Theorem 1.42 are very strong results, since many random variables have generalized concavity properties and are preserved by many operations, such as affine transformations, marginalization and convolution. In [122, p.57] one finds a list of random variables having α-concave densities. We further recommend [121] for a review on α-concave densities and preservation theorems.

Example 1.4. *The density of the non-degenerate multivariate normal distribution $\xi \sim \mathcal{N}(\mu, \Sigma)$ on $\mathbb{R}^d$ given by*

$$f_{\xi}(z) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2}(z-\mu)^T \Sigma^{-1}(z-\mu)\right),$$

where Σ is a positive definite matrix of dimension $d \times d$ and $\mu \in \mathbb{R}^d$. Since the function $ln\left(f_{\xi}(x)\right)$ is concave, i.e. f is 0-concave, the normal distribution is a log-concave measure.

Example 1.5. *i) The Gamma distribution with $\alpha \geq 1$ degrees of freedom, which has density $f_{\xi}(z) = \Gamma(\alpha)^{-1} x^{\alpha-1} e - z\mathbf{1}_{(0,\infty)}(z)$, is log-concave of order α.*

ii) The Weibull density of parameter $\beta \geq 1$, given by $h_{\beta}(z) = \beta z^{\beta-1} \exp(-z^{\beta})\mathbf{1}_{(0,\infty)}(x)$ is log-concave of order β. (The Weibull distribution, and its special case, the Rayleigh, are relevant in energy management, for example, for the wind speed distribution [68]).

Types of Constraint Systems

In the following, we will analyze the existing results for Lipschitz continuity and differentiability of the probabilistic constraints. The first models that we will study are linear in the random vector. This means that the constraint mapping g in (1.19) has one of the following forms

$$g(x,\xi) = h(x) - A\xi \quad \text{or} \quad g(x,\xi) = A(\xi)h(x) - b, \tag{1.25}$$

where A and $A(\xi)$ are deterministic and stochastic matrices, respectively and $h : \mathbb{R}^n \to \mathbb{R}^k$ is a mapping only depending on the decision vector $x \in \mathbb{R}^n$. We also have that $b \in \mathbb{R}^k$. The first model is *separated* from the decision vector, whereas in the second one, the random variable and decision vector are *coupled*. In the following sections of this chapter, we will consider a special case of the linear separated probabilistic constraint: the *random right-hand side* constraint. In this case, transformation A reduced to the identity and the probability constraint (1.19) becomes

$$\varphi_0(x) = \mathbb{P}\left(h(x) \geq \xi\right) \geq p. \tag{1.26}$$

As we can observe, the probabilistic constraint involves inequalities with random variables on the right hand side only. In this case, we can express (1.26) as a constraint on a distribution function

$$F_\xi(h(x)) \geq p$$

where F_ξ is the cumulative multivariate distribution function of the random vector ξ. This composition $\varphi_0 = F \circ h$ allows one to transfer analytical properties like continuity, Lipschitz continuity or differentiability from F_ξ and h to φ_0. afm

Lipschitz Continuous Probabilistic Constraints

In this section we characterize situations in which random right-hand side probabilistic constraints (1.26) have a Lipschitz-continuous distribution function. This situation happens when the marginal distribution functions of the random variable are essentially bounded and the mapping h is Lipschitz continuous.

Theorem 1.44. *[70, 122] Let $\xi \in \mathbb{R}^d$ be a random variable admitting a density with respect to the Lebesgue measure. Suppose that all one-dimensional marginal distribution functions of the random variable $f_\xi^i \in L^\infty(\mathbb{R})$ (i=1, . . . , k) are locally Lipschitz continuous. Then the distribution function F_ξ is also locally Lipschitz continuous.*

The following result shows when the probability function is globally Lipschitz.

Theorem 1.45. *[70] Let $\xi \in \mathbb{R}^d$ be a random variable admitting density with respect to the Lebesgue measure. The distribution function $F_\xi(x) := \mathbb{P}(\xi \leq x)$ is (locally) Lipschitz continuous if and only if the marginal distributions of ξ are essentially bounded, i.e., $f_\xi^{(i)} \in L^\infty(\mathbb{R}) \ \ \forall i = 1, \ldots, d$.*

By combining the property of quasi-concavity with the assumption that ξ is a non-degenerate random variable, we can describe the next stronger result.

Theorem 1.46. *[74] Let $\xi \in \mathbb{R}^d$ be a random variable with quasi-concave law. Its cumulative distribution function $F_\xi(x) := \mathbb{P}(\xi \leq x)$ is Lipschitz continuous if and only if none of its components $\xi_i, \ i = 1, \ldots, d$, have zero variance.*

Differentiability of Probabilistic Constraints

Now we characterize situations in which probabilistic constraints with random right-hand side structure (1.26) are differentiable. In this case, differentiability depends completely on the differentiability of the underlying distribution function of ξ and that of the mapping h. Then, we introduce probabilisitic constraints with two-sided stochastic inequalities and show that they are differentiable for the Gaussian case.

Theorem 1.47. *[107, 70] Let $\xi \in \mathbb{R}^d$ be a random vector with density $f_\xi : \mathbb{R}^d \to \mathbb{R}$. Fix any $z \in \mathbb{R}^d$ and consider $F_\xi(z) = \mathbb{P}(\xi \leq z)$. If*

$$\varphi^{(i)}(t) := \int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_{i-1}} \int_{-\infty}^{z_{i+1}} \cdots \int_{-\infty}^{z_d} f_\xi(u_1, .., u_{i-1}, t, u_{i+1}, .., u_d) du_1 \ldots du_{i-1} du_{i+1} .. du_d,$$

is continuous for all $i = 1, \ldots, d$ then $F_\xi(z)$ is partially differentiable at z and

$$\frac{\partial F_\xi}{\partial z_i}(z) = \varphi^{(i)}(z_i).$$

The following Theorem shows how the derivative of a multivariate normal distribution can be reduced to values of a different multivariate normal distribution in one dimension less:

Theorem 1.48. *[102] Let ξ be an $d-$dimensional Gaussian random vector with mean $\mu \in \mathbb{R}^d$ and positive definite covariance matrix Σ. Then the distribution function $F_\xi(z) = \mathbb{P}(\xi \leq z)$ is continuously differentiable and in any fixed $z \in \mathbb{R}^d$ the following holds:*

$$\frac{\partial F_\xi}{\partial z_i}(z) = f_{\xi_i}(z_i) F_{\bar{\xi}(z_i)}(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_m), \ i = 1, \ldots, d.$$

we denote by $\bar{\xi}(z_i)$ a Gaussian random vector with mean $\bar{\mu} \in \mathbb{R}^{d-1}$ and $(d-1) \times (d-1)$ positive definite covariance matrix $\bar{\Sigma}$. Let D_d^i denote the $d-$th order identity matrix from which the i-th row has been deleted. Then $\bar{\mu} = D_d^i(\mu + \Sigma_{ii}^{-1}(z_i - \mu_i)\Sigma_i)$ and $\bar{\Sigma} = D_d^i(\Sigma - \Sigma_{ii}^{-1}\Sigma_i\Sigma_i^T)(D_d^i)^T$, where Σ_i is the $i-$th column of Σ.

Other important cases concern the computation of gradients of multi-variate Gamma [109] and Dirichlet Distributions [128].

For general linear probabilistic constraints of the type

$$\varphi_1(x) := \mathbb{P}(A\xi \leq x) \tag{1.27}$$

where $x \in \mathbb{R}^n$ is a decision vector, A is a $n \times d$ matrix, and ξ is a d-dimensional non-degenerate Gaussian random vector distributed according to $\xi \sim \mathcal{N}(\mu, \Sigma)$ there is an explicit formulae.

Theorem 1.49. *[71, Thm 3.3] Let $x \in \mathbb{R}^n$ be such that the system $Ax \leq z$ is non-degenarate, where A is a $n \times d$ matrix. Furthermore, let $\xi \sim \mathcal{N}(\mu, \Sigma)$ with $\mu \in \mathbb{R}^d$ and positive definite Σ of order $d \times d$. Then for $j = 1, \ldots, n$, one has the formula*

$$\frac{\partial}{\partial x_j}\mathbb{P}(A\xi \leq x) = f_{\xi}^j(x_j)\mathbb{P}\left(A^{(j)}L^{(j)}\xi^{(j)} \leq x^{(j)} - A^{(j)}w^{(j)}\right)$$

where $\xi^{(j)} \sim \mathcal{N}(0, \mathbb{I}_{d-1})$, $A^{(j)}$ results from deleting row j, $x^{(j)}$ results from x by deleting component j, $w^{(j)} := \mu + \frac{x_j - a_j^T\mu}{a_j^T\Sigma a_j}\Sigma a_j$, $S^{(j)} := \Sigma - \frac{1}{a_j^T\Sigma a_j}\Sigma a_j a_j^T\Sigma$, with factorization $S^{(j)} = L^{(j)}L^{(j)^T}$ and f_{ξ}^j is the one-dimensional Gaussian density with mean value $a_j^T\mu$ and variance $a_j^T\Sigma a_j$. Moreover, the inequality system $A^{(j)}L^{(j)}y \leq x^{(j)} - A^{(j)}w^{(j)}$ is non-degenerate.

For energy management problems, it is important to consider probabilistic constraints, which are induced by two-sided stochastic inequalities of the form:

$$\varphi_2(x) := \mathbb{P}\left(Ax + a \leq \xi \leq Bx + b\right) \tag{1.28}$$

where $\xi \in \mathbb{R}^d$ is a random vector and the vectors $a, b \in \mathbb{R}^d$ and $n \times d$ matrices A, B are deterministic. A system of the form (1.28) is found in hydro reservoir management (see, e.g., [144]).

Proposition 1.50. *[137] Let $\varphi_2 : \mathbb{R}^n \to [0,1]$ be defined as* (1.28) *where $\xi \in \mathbb{R}^d$ is a Gaussian random variable with mean $\mu \in \mathbb{R}^d$ and positive covariance matrix Σ. Then the*

mapping φ_2 is twice differentiable and we have:

$$\begin{aligned}\nabla\varphi_2 =& \nabla_a F_{\xi}(a,b)^T A + \nabla_b F_{\xi}(a,b)^T B \\ \nabla^2\varphi_2 =& A^T\nabla^2_{aa}F_{\xi}(a,b)A + A^T\nabla^2_{ab}F_{\xi}(a,b)B + B^T\nabla^2_{ba}F_{\xi}(a,b)A + B^T\nabla^2_{bb}F_{\xi}(a,b)B\end{aligned}$$

where $F_{\xi}(a,b) := \mathbb{P}\,(a \leq \xi \leq b)$.

One can also compute the derivative of a probabilistic constraint wherein the inequality system is

$$\mathbb{P}\,(A(\xi)x \leq b)$$

and $A(\xi)$ has a multivariate Gaussian distribution. The result can be found in [142]. For a general abstract gradient formula for a probability function we refer to [136].

In order to demonstrate the usefulness of the derived formulas, we consider the optimization problem

$$\begin{aligned}\min \quad & C(x) \\ \text{s.t.} \quad & \varphi_i(x) \geq p\end{aligned} \tag{1.29}$$

where $c, x \in \mathbb{R}^n$ and $C : \mathbb{R}^n \to \mathbb{R}$ is a convex mapping. Given that ξ has a log-concave distribution , we know from Theorem 1.36 that the linear function $x \mapsto \varphi_i(x)$ (for i=0,1, 2) is log-concave. This allows us to transform problem 1.29 into a convex deterministic equivalent:

$$\begin{aligned}\min \quad & C(x) \\ \text{s.t.} \quad & -\log\varphi_i(x) \leq -\log p.\end{aligned} \tag{1.30}$$

Then we can apply algorithms to solve optimization problems involving probabilistic constraints φ_i (for $i = 0, 1, 2$) where we can analytically reduce these values and gradients to values of Gaussian-like distribution functions and compute the latter using standard numerical multiple integration algorithms, which will be discussed in the next chapter.

1.4 Robust Optimization

Robust optimization is another important subfield of optimization under uncertainty. Whereas stochastic programming assumes a probabilistic description of the uncertainty, *robust optimization* deals with a deterministic, set-based description of

the uncertainty. The robust optimization approach constructs a solution that is feasible for any realization of the uncertainty in a given set. The motivation for this approach is computational tractability. The work of [14, 15] and [45] in the late 1990s, coupled with the development of fast interior point methods algorithms for convex optimization paved the way to this young and active research field. We refer the interested reader to the main monograph [12]. In robust optimization, instead of considering problem (1.5), we consider the following parametrized family of problems

$$\begin{aligned} \min \quad & f(x,\zeta) \\ \text{s.t.} \quad & g_j(x,\zeta) \leq 0,\ j = 1,2,\ldots,k. \\ & \zeta \in \mathcal{U} \end{aligned} \tag{1.31}$$

where $f : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ and $g : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^k$ and we assume that ζ lies within a given *uncertainty set* $\mathcal{U} \subseteq \mathbb{R}^d$. We say that a robust optimization problem has convex (quasiconvex, affine, linear) uncertainty, when the functions $f(\cdot,\zeta)$ and $g(\cdot,\zeta)$ are convex (quasiconvex, affine, linear) for all $x \in \mathbb{R}^n$.

Note that problem (1.31) consists of a whole set of parametrized problems, which is often infinitely large. The aim of robust optimization models is to transform this family of problems into a single problem again. The resulting single robust problem is called the *robust counterpart*. We denote by

$$M_\zeta = \{x \in \mathbb{R}^n : g_j(x,\zeta) \leq 0\ \forall j = 1,\ldots,k\}$$

the feasible set of *scenario* $\zeta \in \mathcal{U}$. Furthermore, if there exists a nominal scenario, it is typically denoted by $\hat{\zeta} \in \mathcal{U}$.

A solution $x \in \mathbb{R}^n$ to problem (1.31) is *strictly robust* if it is feasible for all the scenarios in $\mathcal{U}$, i.e., if $g(x,\zeta) \leq 0$ for all $\zeta \in \mathcal{U}$. We denote the set of strictly robust solutions with respect to the uncertainty set $\mathcal{U}$ by

$$\mathcal{M}_\mathcal{U} := \bigcap_{\zeta \in \mathcal{U}} M_\zeta.$$

Thus, the *robust counterpart* of (1.31) is given as

$$\begin{aligned} \min \quad & \sup_{\zeta \in \mathcal{U}} f(x,\zeta) \\ \text{s.t.} \quad & x \in \mathcal{M}_\mathcal{U}. \end{aligned} \tag{1.32}$$

The formulation of the robust counterpart optimization model is connected with the selection of the uncertainty set. The uncertain set is typically selected by the

modeler and plays a vital role in ensuring computational tractability of the robust problem. In the sequel, we introduce several uncertainty sets, which are frequently used in robust optimization applications. The uncertainty set is centered around the nominal values of the uncertain parameters.

i) Finite uncertainty $\mathcal{U} = \{\zeta^1, \ldots, \zeta^N\}$.

ii) Interval-based uncertainty $\mathcal{U} = [\hat{\zeta} - \delta_1, \hat{\zeta} + \delta_1] \times \cdots \times [\hat{\zeta} - \delta_N, \hat{\zeta} + \delta_N]$.

iii) Polyhedral uncertainty $\mathcal{U} = \{\sum_{i=1}^{N} \mu_i \zeta^i : \sum_{i=1}^{N} \mu_i = 1, \mu \in \mathbb{R}_+^N\}$

iv) Ellipsoidal uncertainty $\mathcal{U} = \{\zeta \in \mathbb{R}^d : (\zeta - \hat{\zeta})^T A(\zeta - \hat{\zeta}) \leq \delta\}$ where A is a positive definite matrix.

v) Norm-based uncertainty $\mathcal{U} = \{\zeta \in \mathbb{R}^d : ||\zeta - \hat{\zeta}|| \leq \delta\}$.

The first to consider the idea of associating a *robust counterpart* to the original problem, from a linear programming perspective and for uncertainty sets of the type $\mathcal{U} = K_1, \times \cdots \times K_n$, was Soyster in [124]. The author shows that the robust counterpart can be formulated as a linear program if the sets K_i are compact and convex.

This method was ignored for two decades, and in 1997, it became a strong theoretical framework through the works of Ben-Tal and Nemirovski in [13, 14]. For polyhedral uncertainty sets with convex objective and constraint functions, they show that the robust counterpart preserves many properties of the original problem. The strict robust counterpart of a linear program is again a linear program. Moreover, in such a case, differentiability and convexity are preserved.

Furthermore, the strict robustness for ellipsoidal uncertainty is studied in in [15]. The authors show that the robust counterpart of an uncertain linear program with ellipsoidal (or intersection of ellipsoids) uncertainty is an explicit conic quadratic program. Additionally, the robust counterpart of a convex quadratically constrained program with ellipsoidal uncertainty leads to an explicit semidefinite robust counterpart program.

Robust optimization is a powerful technique because of its tractability. However, complete protection from adverse realizations comes with the cost that solutions are often too conservative. Due to the high conservatism, further research in robust optimization focuses on ways to relax this concept. For example, if the decision maker might allow for a certain degree of constraint violation, the uncertainty set can be selected such that $\mathbb{P}(\zeta \in \mathcal{U}) \geq p$. For a detailed view on this approach see [84].

1.5 Comparing Models

Decision makers would like to know which of the methods is more suitable to employ for optimization problems under uncertainty and weigh the advantages and disadvantages of each one. One should take into account the implementation costs, the operational costs, the behavior of the solutions and the computational tractability.

The robust optimization approach is well suited to the cases where the optimizer wants to hedge the results against all imaginable outcomes of the uncertain events. Since the robust counterpart preserves convexity, differentiability and linearity under mild assumptions of the uncertainty set, it is very often a computationally tractable problem. However, the conservatism of robust optimization may be considerable up to the point of ending up at a very small or even empty feasible sets, possibly coming at much higher costs than under a probabilistic constraint program.

This effect motivates the consideration of probabilistic constraints in the presence of statistical information and of a "well-behaved" (log-concave) distribution. In this case, probabilistic constraints can be easily solved up to few hundred dimensions. As we will see in the next chapter, as a drawback, probabilistic constraints suffer from the curse of dimensionality.

On the other hand, when one has a probabilistic description of the problem (not necessarily "well-behaved"), and ways to quantify the costs of violations, one could model the optimization problem using the recourse approach. However, this approach has a very high computational cost since often the scenarios are discretized.

In [138] a large-scale unit-commitment problem is solved, comparing the three methods discussed above. The results show that the robust optimization is computationally the least costly approach and has the highest recourse cost. The probabilistic constrained approach is second in terms of the computational cost and it improves the solution significantly. The two-stage optimization approach does poorly in terms of robustness (but the recourse decision compensate for this) and the total computational cost is the highest. Similar conclusions are drawn in [87] for a supply planning problem.

Chapter 2

Spherical Radial Decomposition

Section 1.3.3 describes how an optimization problem involving linear probabilistic constraints under a log-concave distribution can be solved by applying deterministic linear programming methods. To do so, one has to provide precise approximations for values and gradients of the probability functions.

In this chapter, we will briefly describe the available numerical integration algorithms Monte-Carlo, quasi-Monte Carlo to numerically approximate the joint probabilistic constraint (1.16). We discuss two breakthroughs related to efficient codes for numerical integration of multivariate probability functions. Firstly, one can efficiently solve linear probabilistic constraints under multivariate normal and t-Student distributions using the Genz code. Secondly, for nonlinear models under the Gaussian random variable, one may profit from the so-called spherical-radial decomposition approach.

2.1 Numerical integration of separable linear JPC

For most functions arising in practice, the integral

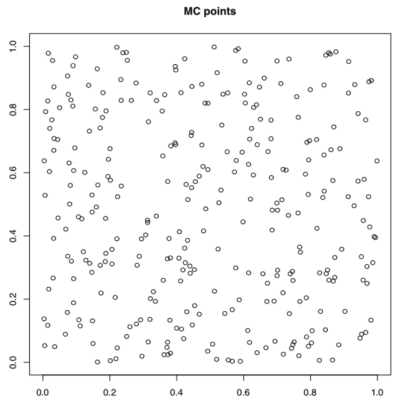
$$\varphi(x) = \int_{\{g(x,z)\leq 0\}} f_{\xi}(z)dz,$$

where $g : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^k$ and $f_{\xi} : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$, cannot be solved analytically. Hence we must solve such integrals numerically; that is, we find an algorithm which enables us to approximate the true value of the integral to a prescribed level of accuracy. The most simple and basic way is crude Monte Carlo. The idea of Monte Carlo integration is to evaluate the integral $\int_A f(x)d\mu(x)$ using random sampling. In its basic form, this is done by independently sampling N points (called *abscissas* or *integration nodes*) $Z_1, \dots Z_N$ according to the given density function f_{ξ} and then

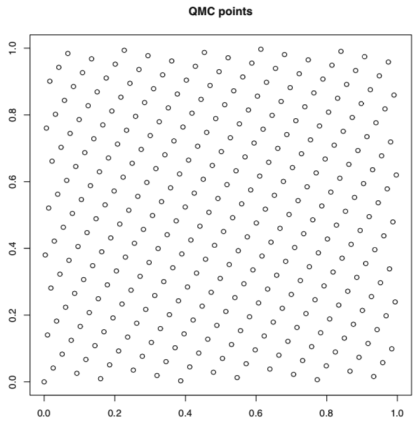
computing the estimate

$$\hat{\varphi}_N(Z_i) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{Z_i \in A},$$

where $A := \{z \in \mathbb{R}^d : g(x,z) \leq 0\}$. The strong law of large numbers guarantees that $\hat{\varphi}_N(x)$ converges a.s. to $\mathbb{E}(\mathbf{1}_A(x)) = \varphi(x)$. If A has very small probability under the law of distribution of ξ, then one might want to consider variance reduction methods [129]. These methods are known to behave poorly in comparison to other methods and we refer to them only for the sake of completeness. In addition, it is also possible to use the quasi-Monte Carlo method, which replaces the random sampling method by well-chosen deterministic abscissas on a grid. The criterion for the choice of deterministic points depends on the problem at hand. The nature of the quasi-Monte Carlo method, with its deterministic procedures, implies that we get deterministic and guaranteed error bounds. Moreover, with the same computational effort, the quasi-Monte Carlo method achieves a significantly higher accuracy [83].



(A) Abcissas of a Monte Carlo integration routine projected onto the x_1x_2 plane.



(B) Abcissas of a quasi-Monte Carlo integration routine projected onto the x_1x_2 plane, [56].

FIGURE 2.1

These two numerical approaches are more efficient when exploiting the special structure of the model (the function g defining A and the distribution of ξ). For instance, for the special case of (1.26) with $g(x,\xi) = x - \xi$ and ξ following a non-degenerate Gaussian distribution, we have that the probability constraint reduces to a multivariate Gaussian distribution function. In this case, Genz [55] developed a numerical integration scheme, where he applies a sequence of three transformations to transform the original integral into an integral over a unit hyper-cube, and in [56] he shows how to solve this numerical scheme using randomized quasi-Monte Carlo method. Using this method, one can compute values of the probabilistic constraint

at a good precision and for a few hundred dimensions. The procedures proposed, called *Genz's code*, are implemented in package *mvtnorm*, available at CRAN. The two functions which are implemented are *pmvnorm* for Gaussian and *pmvt* for the computation of multivariate t probabilities. A similar technique is also proposed in [56] for the multivariate Student distribution. Moreover, the evaluation of the gamma and exponential multivariate distribution has been discussed in [134, 96].

To efficiently use deterministic numerical optimization techniques, its is important to calculate both the functional values of $\varphi(x)$ and its gradients. There are several methods to calculate gradients of probability functions [53, 135, 90]. In the separable case under Gaussian distribution, it is known that partial derivatives of (1.26) can be reduced analytically to function values of a Gaussian distribution with different parameters [107, p.181]. In this way, for the separable case one can apply the same efficient method of Genz [55], mentioned above. Conveniently, this methods can be extended to general linear probabilistic constraints (1.27) under Gaussian distribution with some regularities imposed on the matrix [144, 142, 72].

Therefore, for general linear probabilistic constraints and bilinear ones, with ξ following a Gaussian distribution (with dimension up to a few hundred), one can solve the deterministic counterpart of the probabilistic constrained optimization problem with Genz's code and an SQP solver. To see an application of these methods for a renewable energy management problem see, for instance, [27].

2.2 Numerical Integration for nonlinear JPC

For optimization problems with a more general nonlinear joint probabilistic constraint under normally distributed uncertain parameters, the reduction to Gaussian distribution functions is not possible any more. In this case, another approach, the so-called *spherical-radial decomposition* of Gaussian random vectors, can be used to calculate the functional values of $\varphi(x)$ together with $\nabla_x \varphi(x)$. The integral corresponding to the probability is rewritten via an inner radial integral and an outer spherical integral, where an integration rule is applied by randomly rotating a predetermined set of well-located points. This reparameterization has been developed independently by [38, 113, 92]. Although the spherical-radial integral transformation is widely used, to the best of our knowledge, there is no detailed proof; for the sake of completeness we provide a proof, which relies on simple techniques from multivariable analysis. We first present some preliminary definitions from [118], which we slightly modify for our exposition and then solve [118, exer. 8.6].

Observation 2.2.1. *First of all we would like to extend the familiar formulas for integration of polar coordinates in $\mathbb{R}^3$ to $\mathbb{R}^d$. We recall that with polar coordinates $\Upsilon : \mathbb{R}^3\backslash\{0\} \to (0,\infty) \times [0,\pi] \times [0,2\pi)$, $\Upsilon(x_1,x_2,x_3) = (r\sin\theta_1\cos\theta_2, r\sin\theta_1\sin\theta_2, r\cos\theta_1)$ and the formula*

$$\int_{\mathbb{R}^3} f(x_1,x_2,x_3)dx_1dx_2dx_3 = \int_0^\infty \int_0^\pi \int_0^{2\pi} f(r\sin\theta_1\cos\theta_2, r\sin\theta_1\sin\theta_2, r\cos\theta_1)d\theta_2 r^2\sin\theta_1 d\theta_1 dr$$

for $f : \mathbb{R}^3 \to \mathbb{R}$. These formulas were given using the theory of Riemann integral. We would like to put them in the context of Lebesgue integration, with respect to well-defined measures.

Let us first define the polar coordinates for d dimensions.

Definition 2.1. Let $\mathbb{S}^{d-1} := \{s \in \mathbb{R}^d : ||s|| = 1\}$ denote the unit sphere. For $x \in \mathbb{R}^d\backslash\{0\}$ we write

$$x = ru, \quad \text{where} \quad r = ||x||, \quad u = \frac{x}{||x||}.$$

These are called *polar coordinates* in $\mathbb{R}^d$. Moreover, the corresponding map is

$$\Upsilon : \mathbb{R}^d\backslash\{0\} \to (0,\infty) \times \mathbb{S}^{d-1}$$

$$\Upsilon(x) = (r,u).$$

We note that Υ is a continuous bijection and its inverse is also continuous given by $\Upsilon^{-1}(r,u) = ru$. Therefore, the decomposition in polar coordinates allows us to say that $\mathbb{R}^d\backslash\{0\} = (0,\infty) \times \mathbb{S}^{d-1}$.

Observation 2.2.2. *Let us recall the volume and area of a part of the unit sphere $\mathbb{S}^2$. Let $\Lambda_{\theta_0,\phi_1,\phi_2} := \{(r\sin\phi\cos\theta, r\sin\phi\sin\theta, r\cos\phi) : 0 < r < R, \phi_1 < \phi < \phi_2, \alpha - \frac{\theta_0}{2} < \theta < \alpha + \frac{\theta_0}{2}\}$ be a sector in the sphere $\mathbb{S}^2$ we can easily compute its volume*

$$\mu(\Lambda_{\theta_0,\phi_1,\phi_2}) = \frac{\theta_0 r^3}{3}[\cos\phi_1 - \cos\phi_2],$$

as well as the surface area of the corresponding rectangle $R_{\phi_1,\phi_2,\theta_0}$

$$\sigma(R_{\phi_1,\phi_2,\theta_0}) = r^2\theta_0(\cos\phi_1 - \cos\phi_2).$$

Analogously, we define two important measures for polar coordinates in d-dimensions.

Definition 2.2. Let $B_d := \{s \in \mathbb{R}^d : ||s|| < 1\}$ and denote by μ_d the Lebesgue measure on the $\mathcal{B}(B_d\backslash\{0\})$. Let μ_d^* be a measure on $(0,\infty) \times \mathbb{S}^{d-1}$ induced by the

formula

$$\mu_d^*(A) = \mu_d(\Upsilon^{-1}(A)),$$

where $A \in \mathcal{B}((0,\infty)) \otimes \mathcal{B}(\mathbb{S}^{d-1})$ and $\Upsilon^{-1}(A) = \{ru : r \in (0,1),\ u \in A\}$. Moreover, let ρ be a measure on $(0,\infty)$ given by

$$\rho(E) := \int_E r^{d-1} dr,$$

where $E \subseteq (0,\infty)$.

Definition 2.3. Let $E \in \mathcal{B}(\mathbb{S}^{d-1})$. For $a > 0$ define

$$E_a := \Upsilon^{-1}((0,a] \times B) = \{x \in \mathbb{R}^d | x = ru, 0 < r \leq a, u \in E\},$$

and define a measure σ_{d-1} in $\mathbb{S}^{d-1}$ as follows

$$\sigma_{d-1}(E) = d \cdot \mu_d(E_1).$$

Since the map $E \mapsto E_1$ takes Borel sets in $\mathbb{S}^{d-1}$ to Borel sets in $\mathbb{R}^d$ it is clear that σ is a measure on the σ-algebra of Borel sets in $\mathbb{S}^{d-1}$.

Lemma 2.4. *Let E_a and σ_{d-1} is as in Definition 2.3 and $E \in \mathcal{B}(\mathbb{S}^{d-1})$ then*

$$\mu_d(E_a) = \frac{a^d}{d} \sigma_{d-1}(E).$$

Proof. Let $T_a(x) = ax$ for $a > 0$ and $x \in \mathbb{R}^d$, from the change of variables formula we have that

$$\mu_d(E_a) = \mu_d(T_a(E_1)) = |\det T| \mu_d(E_1) = a^d \mu_d(E_1).$$

Lastly, it follows from Definition (2.3) that $\mu_d(E_a) = \frac{a^d}{d} \sigma_{d-1}(E)$. □

Theorem 2.5. *[118, exer. 8.6] Let $d \in \mathbb{N}$, $d \geq 2$ and let $f : \mathbb{R}^d \to [0,\infty]$ be Borel measurable. Then the Lebesgue measure space $(\mathbb{R}^d \backslash \{0\}, \mathcal{B}(\mathbb{R}^d \backslash \{0\}), \mu_d)$ coincides with $((0,\infty) \times \mathbb{S}^{d-1}, \mathcal{B}((0,\infty)) \otimes \mathcal{B}(\mathbb{S}^{d-1}), \rho \otimes \sigma_{d-1})$ and*

$$\int_{\mathbb{R}^d} f(x) d\mu_d(x) = \int_0^\infty \int_{\mathbb{S}^{d-1}} f(ru) d\sigma_{d-1}(u) r^{d-1} dr.$$

Proof. To identify $\mathcal{B}((0,\infty)) \otimes \mathcal{B}(\mathbb{S}^{d-1})$ and $\mathcal{B}(\mathbb{R}^d \backslash \{0\})$ it suffices to show that the rectangles $A \times \mathbb{S}^{d-1}$ and $(0,\infty) \times B$ belong to $\mathcal{B}(\mathbb{R}^d \backslash \{0\})$ for $A \in \mathcal{B}((0,\infty))$ and $B \in \mathcal{B}(\mathbb{S}^{d-1})$, and that for any open set $G \subset \mathbb{R}^d$ one has that $G \backslash \{0\} \in \mathcal{B}((0,\infty)) \otimes$

$\mathcal{B}(\mathbb{S}^{d-1})$. Since $\mathbb{R}^d\backslash\{0\}$ is homeomorphic to $(0,\infty)\times\mathbb{S}^{d-1}$), then any open $G\subset\mathbb{R}^d$ is a product of two open sets and hence belongs to $\mathcal{B}((0,\infty))\otimes\mathcal{B}(\mathbb{S}^{d-1})$. On the other hand, if we define

$$\Lambda_1 := \{A\in\mathcal{B}((0,\infty))|A\times\mathbb{S}^{d-1}\in\mathcal{B}(\mathbb{R}^d\backslash\{0\})\}$$

$$\Lambda_2 := \{B\in\mathcal{B}(\mathbb{S}^{d-1})|(0,\infty)\times B\in\mathcal{B}(\mathbb{R}^d\backslash\{0\})\}$$

then we have σ-algebras containing open sets, so $\Lambda_1=\mathcal{B}((0,\infty))$ and $\Lambda_2=\mathcal{B}(\mathbb{S}^{d-1})$. Next, we claim that $\mu_d=\rho\otimes\sigma_{d-1}$. It suffices to see that both measures coincide on rectangles $[a,b)\times E$ where $E\in\mathcal{B}(\mathbb{S}^{d-1})$ and $0<a<b$. It follows from Lemma 2.3 that

$$\begin{aligned}
\mu_d^*([a,b)\times E) =& \mu_d(\frac{E_b}{E_a})\\
&= \mu_d(E_b)-\mu_d(E_a)\\
&= \frac{b^d}{d}\sigma_{d-1}(E)-\frac{a^d}{d}\sigma_{d-1}(E)\\
&= \frac{b^d-a^d}{d}\sigma_{d-1}(B)\\
&= \left(\int_a^b r^{d-1}dr\right)\cdot\sigma_{d-1}(B)\\
&= \rho\otimes\sigma_{d-1}.
\end{aligned}$$

Since both measures coincide then for simple functions $f=\sum_{i=1}^m\alpha_i\mathbf{1}_{E_i}$ we have

$$\begin{aligned}
\int_{\mathbb{R}^d} f d\mu_d &= \int_{\mathbb{R}^d\backslash\{0\}} f d\mu_d\\
&= \sum_{i=1}^m \alpha_i\int_{\mathbb{R}^d\backslash\{0\}}\mathbf{1}_{E_i}d\mu_d\\
&= \sum_{i=1}^m \alpha_i\int_0^\infty r^{d-1}\left(\int_{\mathbb{S}^{d-1}}\mathbf{1}_{E_i}(ru)d\sigma_{d-1}(u)\right)dr\\
&= \int_0^\infty r^{d-1}\int_{\mathbb{S}^{d-1}} f(ru)d\sigma_{d-1}(u)dr.
\end{aligned}$$

The standard argument using the Monotone Convergence Theorem gives the general case. □

Lemma 2.6. *[8, Sect 1.3] The surface area of the unit sphere $\mathbb{S}^{d-1}$ is*

$$Area(\mathbb{S}^{d-1}) := \frac{2\pi^{\frac{d}{1}}}{\Gamma(\frac{d}{2})}.$$

Theorem 2.7. *Let ξ be a d-dimensional Gaussian random vector distributed according to $\xi \sim \mathcal{N}(\mu, \Sigma)$. Then for any Borel measurable subset $M \subseteq \mathbb{R}^d$ it holds that*

$$\mathbb{P}(\xi \in M) = \int_{v \in \mathbb{S}^{d-1}} \mu_\chi\{r \geq 0 \,|\, (rLv + \mu) \cap M \neq \emptyset\} d\mu_\eta \tag{2.1}$$

with $\xi = \mu + \chi L\eta$, where L is such that $\Sigma = LL^T$ (e.g., Cholesky decomposition), χ has a chi-distribution μ_{χ_d} with d degrees of freedom and η has a uniform distribution μ_η over the Euclidean unit sphere $\mathbb{S}^{d-1}$.

Proof. Let $M \subseteq \mathbb{R}^d$ be a Borel measurable set. It holds that

$$\mathbb{P}(\xi \in M) = \mathbb{E}(\mathbf{1}_M(\xi)) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \int_{\mathbb{R}^d} \mathbf{1}_M(x) e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} dx. \tag{2.2}$$

Note that a d-variate normal random variable $\xi \sim \mathcal{N}(\mu, \Sigma)$ can be expressed by $\xi = \mu + L\zeta$, where $\zeta \sim \mathcal{N}(0, \mathbb{I})$, and L is the lower triangle matrix such that $\Sigma = LL^T$ (the so-called Cholesky decomposition of Σ). By change of variables $x = \mu + Lz$, we rewrite (2.2) as

$$\mathbb{P}(\xi \in M) = \int_{\mathbb{R}^d} \mathbf{1}_{\tilde{M}}(z) \frac{1}{\sqrt{(2\pi)^d}} e^{-\frac{z^T z}{2}} dz = \mathbb{P}\left(\zeta \in \tilde{M}\right), \tag{2.3}$$

where $\tilde{M} = L^{-1}(M - \mu)$. We do a second change of variables, from the rectangular vector $z \in \mathbb{R}^d \backslash \{0\}$ to polar coordinates $(r, u) \in \mathbb{R}_+ \backslash \{0\} \times \mathbb{S}^{d-1}$ (see Definition 2.1). By Theorem 2.5 this changes the integral accordingly to

$$\mathbb{P}\left(\zeta \in \tilde{M}\right) = \int_{v \in \mathbb{S}^{d-1}} \int_0^\infty \mathbf{1}_{\tilde{M}_v}(r) \frac{1}{\sqrt{(2\pi)^d}} e^{-\frac{r^2}{2}} r^{d-1} dr d\sigma_{d-1}(v),$$

where the translated set is also reparametrized as $\tilde{M}_v = \{r \geq 0 : \ rLv - \mu \in M\}$ for a fixed $v \in S^{d-1}$. Now, let us define $\mu_\eta := \frac{1}{Area(\mathbb{S}^{d-1})} d\sigma_{d-1}$ as the uniform distribution

over $\mathbb{S}^{d-1}$. Then $d\mu_\eta$ is a density of ν (since it integrates one) and it follows that

$$\begin{aligned}
\mathbb{P}\left(\zeta \in \tilde{M}\right) =& \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})} \int_{\mathbb{S}^{d-1}} \int_{\{r\geq 0:\ rLv-\mu\in M\}} \left(\frac{1}{\sqrt{(2\pi)^d}} e^{-\frac{r^2}{2}} r^{d-1} \right) dr d\mu_\eta(v) \\
=& \int_{\mathbb{S}^{d-1}} \int_{\{r\geq 0:\ rLv-\mu\in M\}} \left(\frac{2^{1-\frac{d}{2}}}{\Gamma(\frac{d}{2})} e^{-\frac{r^2}{2}} r^{d-1} \right) dr d\mu_\eta(v) \\
=& \int_{\mathbb{S}^{d-1}} \int_{\{r\geq 0:\ rLv-\mu\in M\}} d\mu_{\chi_d}(r) d\mu_\eta(v) \\
=& \int_{\mathbb{S}^{d-1}} \mu_{\chi_d}\{r \geq 0 \,:\, (rLv+\mu)\cap M \neq \emptyset\} d\mu_\eta.
\end{aligned}$$

□

Observation 2.2.3. *The statement and the proof we give of Theorem 2.7 relies on M being Borel measurable. This hides another difficulty, since for example, not all convex sets are Borel measurable. We would like that Theorem 2.7 be true for M a Lebesgue measurable set (since convex sets are Lebesgue measurable [46]). However, in the following, we will work with closed sets M defined by functions with nice smooth properties, so that M will be Borel and Lebesgue measurable.*

In this way, the probability (2.1) can be numerically computed through an efficient sampling scheme on $\mathbb{S}^{d-1}$ proposed by Deák in [39, 40]. More generally, one may approximate the integral

$$\int_{v\in\mathbb{S}_{d-1}} h(v) d\mu_\eta$$

for any Borel measurable function $h : \mathbb{S}^{d-1} \to \mathbb{R}$. In particular, for

$$h(v) := \mu_\chi\left(\{r \geq 0 : rLv + \mu \cap M \neq \emptyset\}\right).$$

2.3 Gradient formulae for the spherical-radial parametrization of JPC

In [140] the authors show how with a different integrand $h(v)$ the same sampling scheme can be used in order to compute $\nabla\varphi$ and φ simultaneously. In the following, we assume that $g : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ is a continuously differentiable function that is concave with respect to the second argument. We defined

$$\varphi(x) = \mathbb{P}\left(g(x,\xi) \geq 0\right) \tag{2.4}$$

where without loss of generality we assume that $\xi \sim \mathcal{N}(0, \Sigma)$. By (2.1) and (2.4), we have, for all $x \in \mathbb{R}^n$,

$$\varphi(x) = \int_{v \in \mathbb{S}^{d-1}} \mu_\chi \left(\{ r \geq 0 : g(x, rLv) \geq 0 \} \right) d\mu_\eta = \int_{v \in \mathbb{S}^{d-1}} e(x, v) d\mu_\eta$$

for $e(x, v) := \mu_\chi \left(\{ r \geq 0 : g(x, rLv) \geq 0 \} \right)$, $v \in \mathbb{S}^{d-1}$. To treat $\nabla \varphi$, a minimum requirement is that φ has to be continuous, i.e. g is continuous and satisfies Slater condition. However, the following counter-example exposed in [140] shows that this is not enough. It follows that the differentiability for φ requires an additional assumption of the set $\{ z \in \mathbb{R}^d : g(\bar{x}, z) \leq 0 \}$ being compact or, if one wants to relax this requirement and admit noncompact sets, we need a growth condition for the function $||\nabla_x g(x, \cdot)||$ in a neighborhood of $\bar{x}$.

Proposition 2.8. *[140, Prop. 2.2] Let $g : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ be defined as*

$$g(x_1, x_2, z_1, z_2) := x_1^2 e^{h(z_1)} + x_2 z_2 - 1, \quad \textit{where} \quad h(t) := -1 - 2\log(1 - \Phi(t)).$$

Let $\xi \sim \mathcal{N}_2(0, \mathbb{I})$ and $\bar{x} = (0, 1)$, then the following is true;

1. *g is continuously differentiable.*
2. *g is convex in the second argument.*
3. *$g(\bar{x}, 0) = g(0, 1, 0, 0) < 0$.*
4. *φ is not differentiable at $\bar{x}$.*

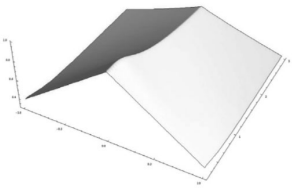


FIGURE 2.2: Illustrates the graph of the non-differentiable function $\varphi(x_1, x_2)$ constructed and depicted in [140]

We now address the computation of $\nabla \varphi$, but first we provide some preliminary definitions and a lemma.

We define the following sets of directions with finite and infinite intersection length (respectively):

$$F(x) := \{ v \in \mathbb{S}^{d-1} | \exists r > 0 : g(x, rLv) = 0 \}$$

$$I(x) := \{v \in \mathbb{S}^{d-1} | \forall r > 0 : g(x, rLv) \neq 0\}.$$

Lemma 2.9. *[140] Let (x, v) be such that $g(x, 0) > 0$ and $v \in F(x)$. Then there exist neighborhoods U of x and V of v, as well as a continuous differentiable function $\rho^{x,v} : U \times V \to \mathbb{R}_+$ with the following properties:*

i) For all $(x', v', r') \in U \times V \times \mathbb{R}_+$ the equivalence $g(x', r'Lv') = 0$ if and only if $r' = \rho^{x,v}(x', v')$ holds true.

ii) For all $(x', v') \in U \times V$ one has the gradient formula

$$\nabla_x \rho^{x,v}(x', v') = -\frac{1}{\langle \nabla_z g(x', \rho^{x,v}(x', v')Lv'), Lv' \rangle} \nabla_x g(x', \rho^{x,v}(x', v')Lv').$$

Definition 2.10. Let $g : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ be a differentiable function. We say that g satisfies the *polynomial growth condition* at x if there exist constants $C, \theta > 0$ and a neighborhood $U(x)$ such that

$$||\nabla_x g(x', z)|| \leq ||z||^\theta \ \ \forall x' \in U(x) \ \forall z, ||z|| \geq C.$$

Theorem 2.11. *[140, Thm. 3.10] Let $g : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ be a continuously differentiable function that is concave with respect to the second argument. Consider the probability function defined in* (2.4)*, where $\xi \sim \mathcal{N}(0, \Sigma)$ has a standard Gaussian distribution with correlation matrix Σ. Let the following assumptions be satisfied at some $\bar{x}$:*

1. *$g(\bar{x}, 0) > 0$.*
2. *g satisfies the polynomial growth condition at $\bar{x}$.*

Then φ is continuously differentiable on a neighborhood $U(\bar{x})$ and it holds that

$$\nabla \varphi(x) = -\int_{v \in F(x)} \frac{\chi(\rho^{x,v}(x, v))}{\langle \nabla_z g(x, \rho^{x,v}(x, v)Lv), Lv \rangle} \nabla_x g(x, \rho^{x,v}(x, v)Lv) d\mu_\eta(v) \ \ \forall x \in U(\bar{x}). \tag{2.5}$$

Here μ_η is the law of the uniform distribution over $\mathbb{S}^{d-1}$, χ is the density of the chi-distribution with d degrees of freedom, L is a factor of the Cholesky decomposition $\Sigma = LL^T$, $F(x)$ and $\rho^{x,v}$ are as introduced in Lemma 2.9.

In this fashion, formula (2.5) is explicit and can be used inside Deák's method to calculate both $\nabla\varphi$ and φ with the same sampling on $\mathbb{S}^{d-1}$. For each sampled point $v \in \mathbb{S}^{d-1}$ one checks whether the equation $g(x, rLv) = 0$ has a solution $r \geq 0$. If it does not, then $v \in I(x)$ and this direction does not contribute to the approximated

integral of the probability. Otherwise, we have that $v \in F(x)$ and we have to evaluate the integrand of (2.5).

Moreover, in [140] also explicit formulae is provided for two non-Gaussian like distributions: χ^2-distribution and t-Student distribution.

Lastly, in [141] the gradient formula was also derived for the case of multiple constraints, i.e. $g : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^k$, where g is a continuously differentiable function. The authors define the maximum function $g^M : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$

$$g^M(x, z) := \max_{i=1,\dots,k} g_i(x, z).$$

Although g^M is not differentiable in general, the authors follow similar previous arguments for the case of convex non-differential analysis (i.e. they address the local Lipschitz continuity of $e(\cdot, v)$ and subdifferential arguments).

Chapter 3

Probust Constraints

In this chapter, we define a new family of probabilistic/robust (*probust*) constraints, where the joint presence of uncertainty in stochastic and unspecified form can be handled in a single model. Applications considered are: a transportation problem in a stationary gas markets with stochastic loads and uncertain roughness coefficients; as well as, an increment of capacity offered by the network operator. For both models, we solved the inner robust problem and then took the outer perspective of probabilistic programming. That is why, we were able to fall back on the spherical-radial decomposition of Gaussian random vectors for probabilistic functions, discussed in Ch. 2.

3.1 Optimization problems under Probust Constraints

Data uncertainty typically enters the inequality constraints describing the set of feasible decisions:

$$g_i(x, z) \geq 0 \quad (i = 1, \ldots, k)\,. \tag{3.1}$$

Here $x \in \mathbb{R}^n$ is a decision vector, $z \in \mathbb{R}^d$ is an uncertain parameter and $g : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^k$ refers to a constraint mapping. Overlooking the aspect of uncertainty would result in optimal decisions that are notoriously non-robust with respect to deviations from the assumed deterministic data. As presented in Ch. 1, when modeling uncertainty, two situations typically occur.

In the first one, access to historical observations is given so that uncertainty can be modeled by means of a random vector ξ obeying a certain estimated multivariate distribution. This allows one to turn (3.1) into a so-called (joint) probabilistic constraint:

$$\mathbb{P}\big(g(x, \xi) \geq 0\big) \geq \rho \in (0, 1) \tag{3.2}$$

(note that the first '$\geq$' sign is to be understood component-wise). The meaning of (3.2) is as follows: a decision x is declared to be feasible if and only if the original random inequality system (3.1) is satisfied with at least probability ρ, a level usually chosen close to but not identical to one, in order to guarantee sufficient robustness without excessive costs.

A second situation arises when some uncertain parameter Φ is, for instance, fixed but simply unknown or it is random but its distribution is unknown because it cannot be observed nor measured. In such a case, the robust optimization approach is introduced to ensure feasibility of its solution for all possible realizations (or, equivalently, for the worst-case scenario) of the uncertain parameter within some uncertainty set $\mathcal{U} \subseteq \mathbb{R}^s$. Then, (3.1) turns into a constraint with infinitely many inequalities

$$g(x, \Phi) \geq 0 \quad \forall \Phi \in \mathcal{U}, \tag{3.3}$$

which can be equivalently written as a single worst-case constraint

$$\inf_{\Phi \in \mathcal{U}} g(x, \Phi) \geq 0. \tag{3.4}$$

Clearly, if g happens to be concave in the first variable, then the set of feasible decisions x in (3.3) or (3.4) is convex, which is a favorable property of optimization problems. Note that both the probabilistic and the worst-case constraints (3.2),(3.3),(3.4) just depend on the decision vector x but unlike (3.1) no longer on the concrete outcome of uncertainty. Hence, they can figure as ordinary well-defined constraints in optimization problems with some additional objective given.

Traditionally, both approaches, probabilistic constraint and robust optimization, have been dealt with separately. Very often, however, one is faced with uncertain variables of both mentioned types. This leads us naturally to the consideration of uncertain inequalities (3.2) in which the uncertain variable has a stochastic and a non-stochastic part, i.e., $z = (\xi, \Phi)$. As a consequence, the originally separate models (3.3) and (3.4) have to be combined appropriately. There are essentially two different ways of doing so.

On the one hand, we formulate a probabilistic constraint (w.r.t. ξ) involving a robustified (w.r.t. Φ) uncertain inequality system:

$$\mathbb{P}\big(g(x, \xi, \Phi) \geq 0 \quad \forall \Phi \in \mathcal{U}\big) \geq \rho. \tag{3.5}$$

This is one joint probabilistic constraint involving infinitely many random inequalities. We call (3.5) under the natural acronym *Probust Constraints*. A second possibility

consists in modeling a family of probabilistic constraints first and then establishing a robustified version of these:

$$\mathbb{P}\big(g(x,\xi,\Phi)\geq 0\big)\geq \rho \quad \forall \Phi\in\mathcal{U}. \tag{3.6}$$

This way, one arrives at an infinite system of probabilistic constraints. This setting is related to (robust) first-order stochastic dominance constraints [42] and to distributionally robust probabiistic constraints, where the probability distribution $\mathbb{P}$ is not fixed (see, e.g. [58]).

Model (3.5) has been studied in [60] under the name of *joint probabilistic-robust constraints*. It has been observed there (and is easy to see) that (3.6) is a weaker constraint than (3.5). Moreover, (3.5) and (3.6) have been introduced and considered before in [139] under the name of *hybrid robust/chance-constraint*. The author considered both models in the context of linear probabilistic constraints under discrete distributions.

Our perspective is different in allowing for nonlinear probust constraints under continuous Gaussian distributions using the so-called spheric-radial decomposition. Moreover, the uncertainty set $\mathcal{U}$ will not be fixed in our two problems but is subject to optimization.

The methodology used in both problems is to first solve the robust counterpart of the inner robust problem of (3.5). Then we get an analytical worst-case solution, which allows us to turn the infinite inequalities system of (3.5) into a joint probabilistic constraint (with finite number of equations). In this context, as shown in Ch. 2, provided that the right hypothesis of the problem hold, we can solve it algorithmically using the spheric-radial decomposition.

3.2 Motivating examples in the natural gas industry

To motivate the use of probust constraints, we present two different optimization problems in the context of gas transport and gas markets, both with uncertain parameters. There are two main applications that drive this new approach.

Firstly, we study the maximum uncertainty allowed when transporting gas with uncertain demand (which is stochastic due to the possibility of fitting statistical distributions based on historical measurements) and with uncertain roughness coefficients in the pipes (which are uncertain but non-stochastic due to a lack of attainable measurements). One ends up with a constraint, which is probabilistic with respect to the load of gas and robust with respect to the roughness coefficients.

Secondly, we study the problem of a network operator who would like to maximize his profit, by offering maximum gas capacity to old clients (who have a stochastic demand but historical data) and to new clients (who have a non-stochastic demand due to a lack of information). One ends up with a constraint, which is probabilistic with respect to the load of old clients and robust with respect to new clients' load.

When modeling optimization problems in the gas markets, it is necessary to understand the legal and economic structure of gas trade as well as the physical structure of gas transport. Thus, in the following, we provide a brief overview of natural gas market regulations and the technical side of gas networks.

3.2.1 Natural gas markets

In the context of the liberalization paradigm, regulatory authorities have separated the natural gas transmission from production and services. The liberalization process of the natural gas market in the EU is a result of three directives passed by the European Commission [131, 132, 133]. The central new aspect is the establishment of market areas based on an entry-exist system and the breaking up of vertically integrated companies. In this way, energy supply is separated from transmission networks. The former system obliged gas suppliers to book an entire transport path through all the gas networks between the desired entry and exit points. Moreover, gas suppliers were very often simultaneously network operators and could openly share (at any necessary time) the information of how much gas could and needed to be transported.

Under the new entry-exit scheme, gas suppliers only need to specify where they want to inject gas (at entry points) and gas consumers where they want to extract gas (at exit-points) within a large market area (without taking into account the transportation path). Today, the gas network operators, which are divided into TSO (Transmission System Operator) operating at high-pressure levels or DSO (distribution system operator) operating low-pressure pipelines, are responsible for calculating and publishing the capacity on the entries and exits of the networks.

The yearly volume network capacity (upper bound) is decided by the network operator, who then sells long-term capacity rights to gas customers in a primary market. Purchasers of long-term capacity rights often include gas producers, industrial facilities, local gas distribution companies and gas marketers, taking speculative positions. No matter how much volume is injected or extracted after a year, the purchasers of the pipeline capacity have to pay a pre-determined price for the upper

bound gas volume agreed on the long-term contract.

Under this new economic configuration new markets have emerged, and as a consequence, new sources of uncertainty and of commercial risks. The network operator, for instance, at the strategic planning stage, has to calculate the yearly network's capacity and the pre-determined price that will cover his/her costs and revenues, without knowing how much demand will there be. In addition, today, gas network operators are companies on their own and strive at maximizing profits.

Moreover, those who have acquired long-term capacity rights, which is called *booking*, have access to a secondary market in a virtual hub, where the booking holders or so-called *gas shippers* sell temporary use of their pipeline capacity at unregulated prices in the day-ahead markets. One day before the booked gas transport is planned to take place, the gas shippers must *nominate* i.e., communicate to the TSO to what extent and where they plan to exercise their capacity rights. This procedure then allows the TSO to schedule the gas transport in advance.

Nominations have to satisfy three conditions:

1. At each node (entry or exit) of the network, nominations must not exceed the capacity booked for that node by the respective trader.

2. Nominations must be balanced over the whole network, i.e., the sum of nominations at entries equals the sum of nominations at exits.

3. Nominations must be technically feasible in the sense that there exist pressures within given bounds at the nodes and a flow through the network such that the nominations at the exits can be served by the nominations at the entries.

This new paradigm has led to new challenging mathematical optimization problems. In this thesis, we will focus on two problems from the network operator's perspective. One of the most important tasks of the network operator is to validate nominations, i.e., the operator examines if the network's capacity is sufficient to transport a specific amount of gas flow. This task has to be done twice: in the planning stage, when the amount of gas that has to be transported is unknown and one-day before operating, when the amount of load to be transported is known. We will focus on the former.

For a more in-depth analysis of gas markets we recommend [125, 99]. In the following we will go shortly through some of the physical characteristics of gas transport, which are important for both of our applications.

3.2.2 Natural gas transport

Natural gas can either be transported grid-bound via pipelines or liquefied as LNG via ships. Pipeline transport is generally the most economical way to transport large quantities of natural gas. To offset the pressure losses during the transport over long distances, the natural gas has to be compressed again every 100 to 200 km. Thus, the study of natural gas transport focuses on the topology of the gas networks and on the flow of gas. The latter can be done in a time-dependent or time-independent approach.

From a mathematical perspective, a gas transmission network can be represented as a directed graph. In this thesis, we will focus our attention to the special case of a tree-structured gas network $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes representing interconnection points, and $\mathcal{E}$ is the set of arcs representing the pipelines. For including cycles into the consideration, we refer to the recent work [63]. For simplicity, we shall assume, in this thesis, that there is just one single injection node (entry), labeled zero, whereas there are m additional nodes at which gas is withdrawn for consumption (exits). The unique entry will be declared as the root of the tree while the arcs in $\mathcal{E}$ are directed away from the root.

The scope of this work is on modeling stationary gas networks, i.e. we analyze network operation as a snapshot, where the parameters characterizing the flow of gas are independent of time. We shall also assume that the network is passive (i.e., does not contain active elements such as compressors, valves etc.). This simplification allow us to maintain a purely algebraic model without combinatorial aspects.

One of the technical responsibilities of the network operator, which has not changed before or after the liberalization, is securing a reliable transport of the natural gas through the given infrastructure. That is why, any optimization problem from the network operator's perspective involves the following constraints:

1. mass flow balance equation at each node
2. gas flow equation through each pipe
3. pressure limits in constraints at each node

We will refer to p and $\hat{b}$ as the vectors of pressures and loads, respectively, at the nodes, and to q as the vector of flows through the arcs of the network. It is well known that the following relations have to be met in order to satisfy the loads $\hat{b}$

(being positive in the case of withdrawal and negative in the case of injection):

$$\begin{aligned} Aq &= \hat{b} && (3.7)\\ A^T p^2 &= (\operatorname{diag}\Phi)|q|q && (3.8)\\ p^{\min} \leq\ & p \ \leq p^{\max}. && (3.9) \end{aligned}$$

Here, A refers to the incidence matrix of the network, $\operatorname{diag}\Phi$ is a diagonal matrix whose diagonal is given by a vector $\Phi = (\Phi_e)_{e\in\mathcal{E}}$ of roughness coefficients for the pipes (arcs) and $p^{\min}, p^{\max}$ are vectors of lower and upper bounds, respectively, for the pressure at the given nodes.

Functions of vectors, such as $(\cdot)^2$ or $|\cdot|(\cdot)$ are to be understood as vectors with entries defined componentwise. Equations. (3.7), (3.8) refer to the first and second Kirchhoff laws (mass flow conservation and pressure drop equations), whereas the bound constraints (3.9) result from technological and contractual restrictions.

Explicit representation of feasible loads

We recall that a nomination is a vector giving the in and outflow of gas loads at entry and exit nodes. Moreover, the task of *validation of a nomination* is to decide whether the gas network can be operated to fulfill the transportation request mandated by the nomination. This problem is frequently faced by network operators during strategic planning level and during operational level. In the strategic planning level, the network operator has to decide how much yearly free capacity can be offered. After calculating the offered capacity, the network operator runs *nomination scenarios* and checks if they are feasible, i.e., if there exists a pressure and flow profile that fulfills (3.7)-(3.9). In the operation level, one day before the gas transport takes place, the network operator runs again a validation of the known (or given) nominations.

Mathematically, the validation of the nominations is the same as ensuring the feasibility of nominations for a given network environment (technical equipment, pressure bounds, booked capacities for injecting or withdrawing gas). It is important to have access to an explicit description of feasibility in terms of the boundary data without further dependence on physical state variables p, q. In [62] the authors provide a description of the feasibility of nomination vector by means of an explicit inequality system. Such description was provided for cycles and tree-structured networks, but we focus on the latter.

Proposition 3.1 (see [62] Cor. 1)**.** *Consider a tree $G = (\mathcal{V}, \mathcal{E})$ with the unique entry referring to the root node 0. Then the load vector $\hat{b} := (b_0, b)$ with b_0 being the load at*

the entry and b being the vector of withdrawals at exits is feasible in the sense of equations (3.7), (3.8), (3.9) *if and only if the following system of inequalities is satisfied for* $k, \ell = 0, \ldots, m$, $k \neq \ell$*:*

$$h_k(b, \Phi) + (p_k^{\max})^2 - h_\ell(b, \Phi) - \left(p_\ell^{\min}\right)^2 \geq 0, \tag{3.10}$$

where, for $k = 0, \ldots, m$*, we have put:*

$$h_k(b, \Phi) := \begin{cases} \sum\limits_{e \in \Pi(k)} \Phi_e \left(\sum\limits_{t \succeq h(e)} b_t \right)^2 & \text{if } k \geq 1 \\ 0 & \text{if } k = 0 \end{cases} \tag{3.11}$$

Here, for $k, \ell \in \mathcal{V}$*, we write* $k \succeq l$ *if, in* G*, the unique directed path from the root to* k*, denoted* $\Pi(k)$*, passes through* l*. Moreover,* $h(e)$ *refers to the head of arc* $e \in \mathcal{E}$*.*

3.3 Probust constraint model for gas transport management

As mentioned above, the reliability of the gas network operator depends on the accuracy of calculations of the transport capacity and on the security of supply. This concern is called *nomination validation*, i.e., determination whether the given (or potential) nominations (scenarios) of all entry and exit flows are technically and physical feasible under the available infrastructure (see [81]).

The strategic planning of the network operator is a complex task. Here we focus on one of the sub-problems that he faces while doing the feasibility check of uncertain future nominations. At the planning stage, the network operator has to validate the feasibility of covering future load while also facing the uncertainty of the pipes' roughness coefficients.

Nevertheless, it is possible to model the amount of future load by means of a stochastic distribution based on historical data (since previous nominations have taken place). Moreover, the roughness coefficient of a pipe influences the pressure loss of the flowing gas. However, the exact values of the roughness coefficients are only known at the time of commissioning of the network, but they change significantly through time due to the aging of the pipe material and to the deposition of particles. The traditional approach to estimate the transport capacity is to set fixed values for the roughness coefficients based on the well-known Colebrook-White correlation [36]. Nonetheless, this approach underestimates the uncertainty of roughness coefficients; hence there is a miscalculation of the transport capacity.

We develop a novel algorithm to improve the nomination validation procedure (in the planning stage) by taking into account these two types of uncertain parameters. The presence of non-stochastic (roughness coefficients) and stochastic (load) uncertainty motivates us to establish a *probust constraint*. As a result, we investigate the maximum of attainable uncertainty around the roughness coefficients for different pipes in the network, while keeping a high probability of satisfying the demand.

In the research literature, there is a more in-depth study of nomination validation in [100]. The robustness of natural gas flows is examined in [33], and [62] gives an explicit characterization of gas flow feasibility and considers the stochastic nature of exit loads.

3.3.1 Description of the optimization problem

We present and solve an optimization problem, from the viewpoint of the network operator, who needs to validate nominations under stochastic and non-stochastic uncertainty.

As mentioned above, we consider a tree $G = (\mathcal{V}, \mathcal{E})$ with a unique entry referring to the root node 0. The nomination vector is denoted by $\hat{b} := (b_0, b)$ with b_0 being the load at the entry and b being the exit-load. The feasibility of the original enhanced vector $\hat{b}$ is already determined by the feasibility of its exit part b, since nominations at entries equals the sum of nominations at exits. Hence, in the following, we will only speak of the feasibility of the load vector b at exits from the very beginning.

Remark 3.2. We emphasize that both variables of the function h_k occurring in the inequality system (3.26) are uncertain, and of different nature. However, the load vector b is of stochastic nature because historical data on its realization are available. In contrast, the roughness coefficients Φ_e may be endowed only with some nominal values, from which they differ in an unobservable way, for instance, by the aging of the pipes' material and the deposition of particles. Our aim is to consider both types of uncertainty in a probust constraint model as detailed in the introduction.

To set up this model, we start by emphasizing the variability of roughness coefficients in the definition of feasible loads via (3.26): For any given vector Φ of roughness coefficients, we denote the set of feasible loads b (satisfying (3.26)) by M_Φ. In order to take into account the uncertain (non-stochastic) nature of Φ, we assume that it belongs to some uncertainty set $\mathcal{U}_\delta$ that is parameterized by a vector $\delta \in \mathbb{R}^{|\mathcal{E}|}$. For instance, $\mathcal{U}_\delta$ could be a rectangle with side lengths δ_e or an ellipsoid with principal axes δ_e around some nominal vector $\hat{\Phi}$ of roughness coefficients. We strengthen now the definition of feasibility of a load vector b w.r.t. uncertainty in Φ by requiring it

to satisfy (3.26) **for all** Φ in the uncertainty set. This means that for each fixed shape parameter δ, the set of feasible loads b is given by

$$M_\delta := \bigcap_{\Phi \in \mathcal{U}_\delta} M_\Phi \quad (\delta \in \mathbb{R}^{|\mathcal{E}|}). \tag{3.12}$$

Often, the choice of the uncertainty set, in particular of its size, is somehow arbitrary. This motivates us, in this paper, not to fix it but to consider this set as variable (via the parameter δ) and subject to optimization.

On the other hand, in order to address the stochastic uncertainty of the load vector b, we will understand it from now on as the outcome of an m-dimensional random vector ξ, where $m = |\mathcal{V}| - 1$ is the number of exits. In particular, we will assume ξ to be a Gaussian random vector $\xi \sim \mathcal{N}(\mu, \Sigma)$ with mean μ and covariance matrix Σ. In a strict sense, exit loads cannot follow a Gaussian distribution because the latter allows negative values. This effect, however, is negligible if the relative standard deviations are small. Moreover, the methodology we are presenting here is easily adapted to Gaussian-like distributions (truncated Gaussian, Log-normal) which are good candidates for modeling of stochastic exit loads [81, Chapter 13].

Accordingly, we turn the original membership $b \in M_\delta$ from (3.12) into the probabilistic constraint $\mathbb{P}(\xi \in M_\delta) \geq \rho$, where $\rho \in (0,1)$ is a probability level (close to one) chosen in the modeling process. Observe that this inequality now is a constraint on the shape parameter δ for the uncertainty set U_δ. Since we consider δ as a variable, it defines a set of feasible shape parameters as

$$M := \left\{ \delta \in \mathbb{R}^{|\mathcal{E}|} \,\middle|\, \mathbb{P}(\xi \in M_\delta) \geq \rho \right\}. \tag{3.13}$$

The interpretation of this constraint is as follows: a shape parameter δ is feasible if and only if the probability of random load vectors b being admissible in the sense of (3.26) **for all** roughness coefficients $\Phi \in U_\delta$ is at least ρ. Clearly, an increase of δ and, hence of U_δ, the uncertainty set will result in a stronger "for all" condition and, thus, in a decrease of the probability of the event $\xi \in M_\delta$.

With this setup, we are interested in a maximum amount of uncertainty for the roughness coefficients that still allows us to technically satisfy − under the given pressure bounds − the random loads for all uncertain roughness coefficients with a given probability ρ. This leads us to the optimization problem

$$\max\{f(\delta) \,|\, \delta \in M\}, \tag{3.14}$$

where the objective f characterizes the amount of uncertainty, e.g., $f(\delta) := \operatorname{vol} U_\delta$ or $f(\delta) := \|\delta\|$. The solution of (3.14) may provide the network operator with an idea at what precision at least he needs to know the roughness coefficient in the context of safe network operation. Such information could be used, for instance, when trying to roughly estimate these coefficients by indirect measurements via the solution of an inverse problem (see [44]) or in order to identify critical parts of the network where it is more important to do so than in other parts.

Determination of the probability of M_δ

The key for solving the optimization problem (3.14) is clearly the verification of the inclusion $\delta \in M$, which according to (3.13) amounts to the computation of the probability of the event $\xi \in M_\delta$. This task faces two difficulties which we will address in this section: first, the set M_δ, given as an infinite intersection of sets M_Φ has to be made explicit, and second, the Gaussian probability of such a set has to be determined efficiently.

3.3.2 Robust counterpart reformulation of M_δ

By definition of M_Φ as consisting of all load vectors b satisfying (3.26), we may write M_δ in (3.12) as

$$M_\delta = \left\{ b : h_k(b,\Phi) + (p_k^{\max})^2 - h_\ell(b,\Phi) - \left(p_\ell^{\min}\right)^2 \geq 0 \ \forall \Phi \in U_\delta \ \forall k,\ell = 0,\dots,m \ (k \neq \ell) \right\}.$$

The difficulty in working with the set M_δ lies in the fact that it is defined by infinitely many constraints, because the uncertainty set is infinite in general. Evidently, we have the following equivalent reformulation:

$$M_\delta = \left\{ b : \inf_{\Phi \in U_\delta} h_k(b,\Phi) + (p_k^{\max})^2 - h_\ell(b,\Phi) - \left(p_\ell^{\min}\right)^2 \geq 0 \forall k,\ell = 0,\dots,m \ (k \neq \ell) \right\}.$$

In the typical case of compact uncertainty sets U_δ, we may finally represent M_δ as

$$\begin{aligned} M_\delta = \Big\{ b : \ & h_k(b,\Phi^*_{k\ell}(\delta,b)) + (p_k^{\max})^2 - h_\ell(b,\Phi^*_{k\ell}(\delta,b)) \\ & - \left(p_\ell^{\min}\right)^2 \geq 0 \\ & \forall k,\ell = 0,\dots,m \ (k \neq \ell) \Big\}, \end{aligned} \tag{3.15}$$

where

$$\Phi^*_{k\ell}(\delta,b) := \operatorname*{argmin}_{\Phi \in U_\delta} \left(h_k(b,\Phi) - h_\ell(b,\Phi) \right). \tag{3.16}$$

This last representation (3.15) has the advantage of involving only a finite number (at most $m^2 - m$) of inequalities in contrast to the original description. However, it comes at the price of having to solve the optimization problem (3.16) for each b and δ. Fortunately, this is easily done for specific simple enough uncertainty sets, for instance, rectangles or ellipsoids. This is mainly a consequence of the difference function $h_k - h_\ell$ appearing in (3.16) being linear in Φ by definition of h_k in Proposition 3.1. Minimizing a linear function over a rectangle or ellipsoid can be done explicitly in terms of the coefficents of this linear function.

Explicit solutions for simple uncertainty sets

In the following, we will provide explicit solutions for the functions $\Phi^*_{k\ell}(\delta, b)$ introduced in (3.16) in the case of ellipsoidal and rectangular uncertainty sets. We will assume that our uncertainty sets are symmetric around some nominal (or guessed) value $\hat{\Phi}$ for the roughness coefficients. In the case of an ellipsoid, we define for $\delta \in \mathbb{R}_+^{|\mathcal{E}|}$:

$$U_\delta := \left\{\Phi \in \mathbb{R}^{|\mathcal{E}|} : (\Phi - \hat{\Phi})^T \Sigma_\delta (\Phi - \hat{\Phi}) \leq 1\right\}, \tag{3.17}$$

where Σ_δ is a diagonal matrix with entries $\delta_1, \ldots, \delta_{|\mathcal{E}|}$.

In the following, referring back to the notation introduced in Proposition 3.1, we will use the assignments

$$\mathbf{1}_{e,k} := \begin{cases} 1 & \text{if } e \in \Pi(k) \\ 0 & \text{otherwise} \end{cases}$$

and

$$\gamma_e(b) := \sum_{t \in \mathcal{V},\, t \succeq h(e)} b_t \,. \tag{3.18}$$

This notation will allow us to rewrite the definition for $h_k(b, \Phi)$ in Proposition 3.1 as

$$h_k(b, \Phi) = \sum_{e \in \mathcal{E}} \Phi_e \gamma_e^2(b) \mathbf{1}_{e,k} \quad (k = 1, \ldots, m). \tag{3.19}$$

Lemma 3.3. *For the uncertainty set* (3.17)*, the functions introduced in* (3.16) *have the following explicit representation for $k, \ell = 0, \ldots, m$ with $k \neq \ell$ and $e \in \mathcal{E}$:*

$$[\Phi^*_{k\ell}(\delta, b)]_e = \hat{\Phi}_e + \frac{(\mathbf{1}_{e,\ell} - \mathbf{1}_{e,k})\gamma_e^2(b)/\delta_e}{\sqrt{\sum_{s \in (\Pi(k) \cup \Pi(\ell)) \setminus (\Pi(k) \cap \Pi(\ell))} \gamma_s^4(b)/\delta_s}}.$$

Proof. It is well-known (and easy to show by writing down the necessary optimality conditions) that for a given cost vector $c \neq 0$, the optimization problem

$$\min_{\Phi} \left\{ c^T \Phi : (\Phi - \hat{\Phi})^T \Sigma_\delta (\Phi - \hat{\Phi}) \leq 1 \right\} \tag{3.20}$$

has the unique solution

$$\Phi^* = \hat{\Phi} - \frac{1}{\sqrt{c^T \Sigma_\delta^{-1} c}} \Sigma_\delta^{-1} c. \tag{3.21}$$

Clearly, the optimization problem defining $\Phi^*_{k\ell}(\delta, b)$ in (3.16) has the form of (3.20) with the coefficients of the cost vector given by (according to (3.19))

$$c_e := (\mathbf{1}_{e,k} - \mathbf{1}_{e,\ell}) \gamma_e^2(b) \quad (e = 1, \ldots, |\mathcal{E}|).$$

Taking into account that Σ_δ is a diagonal matrix with diagonal entries δ_e, the optimal solution $\Phi^*_{k\ell}(\delta, b)$ can be read off component-wise from (3.21) to yield the asserted formula. □

Alternatively to (3.17), we now introduce a family of rectangular uncertainty sets, again centered around some nominal value $\hat{\Phi}$ by means of

$$U_\delta := [\hat{\Phi} - \delta, \hat{\Phi} + \delta] \quad (\delta \in \mathbb{R}_+^{|\mathcal{E}|}). \tag{3.22}$$

Lemma 3.4. *For the uncertainty set* (3.22)*, the functions introduced in* (3.16) *are actually independent of b and have the following explicit representation for $k, \ell = 0, \ldots, m$ with $k \neq \ell$ and $e \in \mathcal{E}$:*

$$[\Phi^*_{k\ell}(\delta, b)]_e = \begin{cases} \hat{\Phi}_e - \delta_e & \text{if } e \in \Pi(k) \setminus \Pi(\ell) \\ \hat{\Phi}_e + \delta_e & \text{if } e \in \Pi(\ell) \setminus \Pi(k) \\ \hat{\Phi}_e & \text{otherwise} \end{cases}.$$

Proof. We observe from (3.19) and (3.18) that the following holds true for $k, \ell = 0, \ldots, m$ with $k \neq \ell$ and $e \in \mathcal{E}$:

$$\min_{\Phi \in \mathcal{U}_\delta} \{h_k(b, \Phi) - h_\ell(b, \Phi)\} = \min_{\Phi \in \mathcal{U}_\delta} \sum_{e \in \mathcal{E}} (\mathbf{1}_{e,k} - \mathbf{1}_{e,\ell}) \Phi_e \gamma_e^2(b).$$

With $\mathcal{U}_\delta$ being a rectangle and with the sum above being separable in the components Φ_e, the minimization can be carried out component-wise. Accordingly, each

component Φ_e is chosen in the interval $[\hat{\Phi}_e - \delta_e, \hat{\Phi}_e + \delta_e]$ as to minimize the expression

$$(\mathbf{1}_{e,k} - \mathbf{1}_{e,\ell})\Phi_e \gamma_e^2(b).$$

Since the coefficients $\gamma_e^2(b)$ are non-negative, and $(\mathbf{1}_{e,k} - \mathbf{1}_{e,\ell})$ equals ± 1 (or zero, in which case the choice of Φ_e is arbitrary), we can choose the minimizing component as

$$[\Phi_{k\ell}^*]_e := \hat{\Phi}_e - (\mathbf{1}_{e,k} - \mathbf{1}_{e,\ell})\delta_e.$$

By definition, this reduces to the formula asserted in the statement of our lemma. □

In order to illustrate the set M_δ of feasible exit load vectors, we present an example of elementary three-node network with a squared-shaped uncertainty set.

Example 3.1. *In order to illustrate the geometry of set M_δ of feasible exist load vectors, we present an elementary tree gas network consisting of two exit nodes (b_1, b_2), two pipelines i, and a squared uncertainty set around $\hat{\Phi} \in \mathbb{R}^2$ given by $\mathcal{U}_\delta = [\hat{\Phi}_1 - \delta, \hat{\Phi}_1 + \delta] \times [\hat{\Phi}_2 - \delta, \hat{\Phi}_2 + \delta]$. From (3.11) and from Lemma 3.4 we have that*

$$M_\delta = \left\{ (b_1, b_2) \in \mathbb{R}_+^2 \;\middle|\; \begin{array}{l} \underline{y_0} \leq \min\{\overline{y_1} + (\hat{\Phi}_1 - \delta)b_1^2, \overline{y_2} + (\hat{\Phi}_2 - \delta)b_2^2\} \\ \overline{y_0} \geq \max\{\underline{y_1} + (\hat{\Phi}_1 + \delta)b_1^2, \underline{y_2} + (\hat{\Phi}_2 + \delta)b_2^2\} \\ \underline{y_1} + (\hat{\Phi}_1 + \delta)b_1^2 \leq \overline{y_2} + (\hat{\Phi}_2 - \delta)b_2^2 \\ \underline{y_2} + (\hat{\Phi}_2 + \delta)b_2^2 \leq \overline{y_1} + (\hat{\Phi}_1 - \delta)b_1^2 \end{array} \right\}$$

Figure 3.1 illustrates the set of vectors (b_1, b_2) belonging to the feasible set M_δ for different values of δ. The larger the uncertainty δ, the smaller the feasible set M_δ, which translates into a more conservative optimal solution. Moreover, we note how M_δ is a non-convex set, and as δ increases, it becomes more similar to a square.

Computing the probability of feasible random exit loads

Now that we are given an explicit description of the set M_δ in (3.15) for the special case of elliptical or rectangular uncertainty sets (upon substituting the functions $\Phi_{k\ell}^*(\delta, b)$ by the formulae obtained in Lemmas 3.3 and 3.4), we could use this finite inequality system in order to test the feasibility of simulated outcomes of the random load b according to the given Gaussian distribution. The averaged number of feasible simulations would yield the Monte Carlo estimate for the desired probability $\mathbb{P}(\xi \in M_\delta)$. Such Monte Carlo approach has two drawbacks: first it may come with

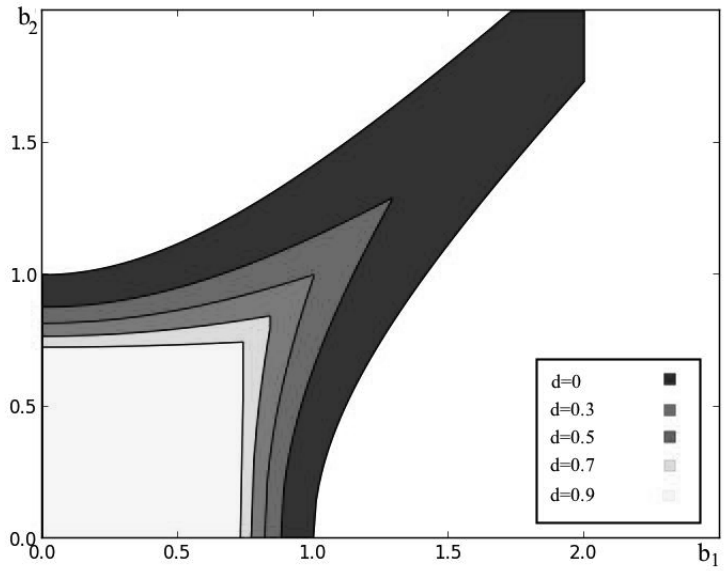


FIGURE 3.1: Illustration of the feasible set of nominations M_δ as δ increases for a square-shaped $\mathcal{U}_\delta$.

a comparatively large variance for the obtained probability estimation and, second, it does not provide us with information about the sensitivity of this probability with respect to changes of δ. This sensitivity (derivative) information is crucial, however, in order to set up any efficient algorithm of nonlinear optimization in order to solve problem (3.14). Therefore, we will alternatively make use of the so-called *spheric-radial decomposition of Gaussian random vectors* exposed in Chapter 3.

In order to evaluate the integrand in the spheric integral of Theorem 2.7, one has to be able to compute, for any fixed direction $v \in \mathbb{S}^{m-1}$, the χ-probability of the one-dimensional set

$$\{r \geq 0 \,|\, (rLv + \mu) \cap A \neq \emptyset\}.$$

Since we are interested in the probability of the set $A := M_\delta$, by (3.15) this amounts to characterizing the set

$$\{r \geq 0 \,|\, g(\delta, rLv + \mu) \geq 0\} \quad (v \in \mathbb{S}^{m-1}), \tag{3.23}$$

where we set

$$g(\delta, b) := \min_{\substack{k,\ell=0,\ldots,m \\ k \neq l}} \left\{ h_k(b, \Phi^*_{k\ell}(\delta, b)) + (p_k^{\max})^2 - h_\ell(b, \Phi^*_{k\ell}(\delta, b)) - (p_\ell^{\min})^2 \right\}. \tag{3.24}$$

Using the idea of spheric-radial decomposition presented in Theorem 2.7, we propose the following algorithm for computing the probability $\mathbb{P}(\xi \in M_\delta)$ with a fixed value of δ:

Algorithm 3.3.1. *Let $\delta \in \mathbb{R}^{|\mathcal{E}|}$ be arbitrary, $\xi \sim \mathcal{N}(\mu, \Sigma)$ and L such that $LL^T = \Sigma$.*

1. *Sample N points $\{v_1, v_2, \ldots, v_N\}$ uniformly distributed on the sphere $\mathbb{S}^{m-1}$.*
2. *$i := 0$; $S := 0$.*
3. *$i := i + 1$;*
   Find the zeros of the one-dimensional function (in r for δ fixed) $\theta_\delta(r) := g(\delta, rLv_i + \mu)$ with g defined in (3.24) *and represent the set $M^i_\delta := \{r \geq 0 \,|\, \theta_\delta(r) \geq 0\}$ corresponding to* (3.23) *as a disjoint union of intervals: $M^i_\delta = \cup_{j=1}^{\rho}[\alpha_j(\delta), \beta_j(\delta)]$, where $\alpha_j(\delta), \beta_j(\delta)$ are the zeros obtained before and ordered appropriately.*
4. *Compute the χ-probability of M^i_δ according to*

$$\mu_\chi(M^i_\delta) = \sum_j F_\chi(\beta_j(\delta)) - F_\chi(\alpha_j(\delta)),$$

   where F_χ refers to the cumulative distribution function of the one-dimensional χ-distribution with m degrees of freedom. Put $S := S + \mu_\chi(M^i_\delta)$
5. *If $i < N$ then go to 3.*
6. *Set $\mathbb{P}(\xi \in M_\delta) := \frac{S}{N}$.*

A few words on this algorithm are in order at this point. The algorithm clearly provides an approximation to the spheric integral in Theorem 2.7 by means of a finite sum based on sampling of the sphere and then averaging the values of the integrand over all samples. Of course, this approximation will improve with the sampling size, which may be large depending on the dimension m of the problem (i.e., exit nodes in the network) and on the desired precision for the probability.

We recall that the uniform distribution on the sphere $\mathbb{S}^{m-1}$ can be represented as the distribution of $\eta/\|\eta\|$ (Euclidean norm), where η has a standard Gaussian distribution in $\mathbb{R}^m$, i.e., $\eta \sim \mathcal{N}(0, I)$. Then the simplest idea is to sample a point v^i on the sphere as in step 1 of the algorithm would be to independently sample m values w_j of a one-dimensional standard normal distribution by using standard random generators and then putting $v_i := w/\|w\|$ for $w := (w_1, \ldots, w_m)$. When replacing such Monte Carlo sampling of the normal distribution (based on random

number generators) by quasi-Monte Carlo sampling (based on deterministic low discrepancy sequences), one observes a dramatic improvement in the precision of the result. For our gas network problem (with fixed roughness coefficients), this was revealed in [62]. A further improvement is to be expected for direct Quasi-Monte Carlo sampling on the sphere (not via normalization of Gaussian distributions) as discussed in [26].

We illustrate the construction of the sets M_i in step 3 of Algorithm 3.3.1 for the special case of uncertainty sets given by rectangles (3.22): thanks to Lemma 3.4, the optimal coefficients $\Phi^*_{k\ell}(\delta, b)$ do not depend on b. Hence, we may simply write them as $\Phi^*_{k\ell}(\delta)$ with values according to Lemma 3.4. Now, by (3.24) and by definition of h_k below (3.26), we have that the function in r whose zeros are looked for in step 3 takes the form

$$\theta_\delta(r) = \min_{\substack{k,\ell=0,\dots,m \\ k\neq\ell}} \left\{ \sum_{e\in\Pi(k)} [\Phi^*_{k\ell}(\delta)]_e \left(\sum_{t\succeq h(e)} rL_t v_i + \mu_t \right)^2 \right. \\ \left. - \sum_{e\in\Pi(\ell)} [\Phi^*_{k\ell}(\delta)]_e \left(\sum_{t\succeq h(e)} rL_t v_i + \mu_t \right)^2 \right\},$$

where L_t denotes row t of the matrix L. Clearly, each of the expressions inside the minimum is quadratic in r; hence, θ may be written as a minimum of quadratic functions

$$\theta_\delta(r) = \min_{\substack{k,\ell=0,\dots,m \\ k\neq l}} \left\{ c_{k\ell}(\delta) + d_{k\ell}(\delta) r + e_{k\ell}(\delta) r^2 \right\}, \tag{3.25}$$

with coefficients easily identified from the formula above. Since the zeros of θ_δ must be contained in the zeros of all single quadratic functions inside the minimum, one may proceed as follows: determine first all zeros of the single quadratic functions above and order them as $x_1, \dots x_K$. Second, delete from this list all zeros x_n for which $x_n < 0$ or $\theta_\delta(x_n) < 0$. The remaining list, say $y_1, \dots, y_{K'}$ will consist of the positive zeros of θ_δ. Third, identify neighboring zero's, between which the function θ_δ remains positive, in order to represent the set M_i in step 3 as a union of disjoint intervals. For instance, if $\theta_\delta(0) > 0$, then we'll have that

$$M_i = [0, y_1] \cup [y_2, y_3] \cup \cdots,$$

whereas for $\theta_\delta(0) < 0$ the representation will be

$$M_i = [y_1, y_2] \cup [y_3, y_4] \cup \cdots.$$

Observe that the last of these intervals will be closed by the last zero $y_{K'}$ if $\theta_\delta(y_{K'}) < 0$ or by ∞ else.

Once a representation of the set M_i in step 3 as a union of disjoint intervals has been obtained, step 4 is easily accomplished by applying efficient standard routines for high-precision approximations of the one-dimensional cumulative distribution function of the χ-distribution.

3.3.3 Numerical solution of the optimization problem

Now we describe the numerical solution of the optimization problem (3.14) and illustrate the results for a concrete example. As a solution method for nonlinear optimization problems subject to inequality constraints we have chosen the projected gradient method [57], as it behaves rather robustly with respect to the inevitable inaccuracy in the computation of probabilities by means of Algorithm 3.3.1. Note that this inaccuracy can be reduced at the cost of increasing computation time by enhancing the sample size N. In order to apply the projected gradient method, it is crucial not only to determine the (δ-dependent) probabilities $\mathbb{P}(\xi \in M_\delta)$ in (3.13) but also their gradients with respect to δ.

Approximating the gradient of the probability function

The sensitivity information on the δ-dependent probability function mentioned above can be gained directly from within the Algorithm 3.3.1: by step 6, the probability is given by $\frac{S}{N}$, where S is updated in step 4. Accordingly, the derivative of the probability with respect to the parameter δ can be approximated by the expression $\frac{S'}{N}$, where S' is updated by the derivatives of the updates of S with respect to δ:

$$\sum_j f_\chi(\beta_j(\delta))\nabla\beta_j(\delta) - f_\chi(\alpha_j(\delta))\nabla\alpha_j(\delta).$$

Here, f_χ refers to the density of the one-dimensional χ-distribution with m degrees of freedom; i.e., f_χ is the derivative of the distribution function F_χ from step 4. Since there exists an analytical expression for f_χ, all one needs to know for evaluating the expression above is the gradients $\nabla\beta_j$, $\nabla\alpha_j$, i.e, the gradients with respect to δ of the appropriate zeros of the function θ_δ defined in step 3). These are easily found by representing the appropriate zeros of the associated quadratic equation

$$c_{k\ell}(\delta) + d_{k\ell}(\delta)r + e_{k\ell}(\delta)r^2 = 0$$

in (3.25) as $r(\delta)$ by using the classical solution formula and then deriving analytically $r(\delta)$ with respect to δ. We emphasize that the procedure sketched above corresponds to calculating the gradient of the approximated probability function. This does not automatically have to coincide with calculating an approximation of the gradient of the true probability function; that is, both operations don't have to commute. It was found in [140, 141] by deriving the corresponding gradient formulae that both approaches coincide under convexity of the underlying random inequality system. Since this assumption is not satisfied in our case for the system (3.26), a rigorous justification of the procedure described above (differentiation 'within the algorithm') is missing so far. On the other hand, our numerical experience suggests that the projected gradient method performs well in finding local solutions to problem (3.14).

Illustration of a toy problem

For the purpose of illustration we start by considering a toy example of problem (3.14). Here, a simple network consisting of one entry node, one passive node (Innode) without injection or consumption and two exit nodes with random load is given as in Figure 3.2.

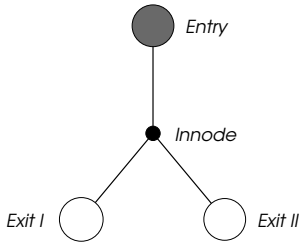


FIGURE 3.2: Structure of a simple network for a toy example containing one entry, one passive node and two exits.

Observe, that the innode can be formally modeled as an exit with zero consumption. As there exist 3 arcs joining the nodes, we have three roughness coefficients Φ_e. We will consider the maximization of rectangular uncertainty sets U_δ as introduced

in (3.22). We assume the following data:

$$\hat{\Phi}_1 = \hat{\Phi}_2 = \hat{\Phi}_3 = 0.0015$$
$$p^{\min} = (1,1,1,1); \quad p^{\max} = (390,200,100,120)$$
$$\xi = (\xi_1, \xi_2) \sim \mathcal{N}\left((4100,3900); \begin{pmatrix} 300 & 0 \\ 0 & 300 \end{pmatrix}\right)$$
$$f(\delta) := (\delta_1)^{0.9} + (\delta_2)^{0.9} + (\delta_3)^{0.9}$$
$$\rho = 0.8.$$

Here, the components of the vector $\hat{\Phi}$ of average roughness coefficients correspond to the arcs (entry-innode, innode-exit 1, innode-exit 2), whereas the components of $p^{\min}, p^{\max}$ are labeled according to (entry, innode, exit 1, exit 2). The two exits are supposed to have Gaussian random load with means 4100 and 3900, respectively and standard deviation 300 both. We assume that the two loads are uncorrelated (which does not necessarily have to be the case in our approach). The objective function f measuring the size of the uncertainty set U_δ has not been chosen on purpose as a norm. The reason is that, e.g., the one-norm (which would correspond to exponents one in the same expression) yields strongly 'concentrated' solutions, i.e. optimal rectangular uncertainty sets with many extremely short sides and a few long ones. In other words, in order to allow a larger uncertainty for a few roughness coefficients, one has to insist on extremely high accuracy for many other. Choosing exponents smaller than one makes it possible to control the contrast between different uncertainty ranges and to maintain a reasonable minimum amount. The probability level ρ in (3.13) was set to 0.8. The numerical solution of problem (3.14) with the given data is:

$$\delta_1^* = 0.00014595; \quad \delta_2^* = 0.00006697; \quad \delta_3^* = 0.00020503.$$

Relating these values to the average roughness coefficients indicated above, the interpretation of this result is as follows: 80% of the random load scenarios at exits 1 and 2 are technically feasible in the sense of Prop. 3.1 **for all** roughness coefficients deviating relatively from their average values by at most 97.3%, 44.6% and 136.7%, respectively. We can make a posterior check of this solution by simulating a set of exit load scenarios according to the Gaussian distribution given above and counting for how many of them feasibility holds true for all roughness coefficients in the calculated maximum uncertainty rectangle.

Figure 3.3 shows ten different simulated load patterns for the two exits (with values scattered around the averages given above indicated in the figure). For each simulation, the optimal rectangular uncertainty with side lengths given above is shown along with all infeasible roughness coefficients colored inside. An empty rectangle indicates that the load pattern is feasible **for all** roughness coefficients in the rectangle. Then, by definition, this load scenario is counted as feasible. According to Figure 3.3, eight out of ten load scenarios are feasible, which corresponds (by chance exactly) to the chosen probability level $\rho = 0.8$. For the two infeasible scenarios, the polyhedral set (actually, the complement of a polyhedron) of roughness coefficients in the rectangle violating feasibility is made visible.

Numerical solution of a medium size problem

In this section, we provide the results solving problem (3.14) in a more realistic setting with a network consisting of 27 nodes (1 entry and 26 exits). In other words, the random vector considered in the probabilistic constraint follows a 26-dimensional Gaussian distribution. The network is illustrated in Figure 3.4 with the entry located in the center. The pipes in the network are colored according to the relative uncertainty allowed for the corresponding roughness coefficients in an optimal (maximal) uncertainty set computed with the same objective as in the previous toy example. The colors follow a temperature scale, where red means that only low uncertainty is allowed whereas blue means high uncertainty can be tolerated. The solution in the left figure was obtained by imposing a probability level $p = 0.8$ whereas the one in the right figure refers to a probability level $p = 0.9$. Not surprisingly, higher accuracy for roughness coefficients is required overall when increasing the probability level. More interestingly, higher accuracy is required at arcs close to the entry.

3.4 Probust constraint model for gas capacity maximization

In this model, we consider a network operator who is in the planning stage; that is, he is considering how much capacity should he/she offer to the gas shippers. At this stage, the nominations are not known, but he/she has historical data for usual clients. In principle, the network owner has to make sure that all nominations complying with the booked capacities can be satisfied by a feasible flow through the network satisfying given lower and upper pressure bounds at its nodes. However, in practice, daily nominations are lower than the offered daily capacity. Therefore, the network operator may be content with guaranteeing feasibility for his nomination

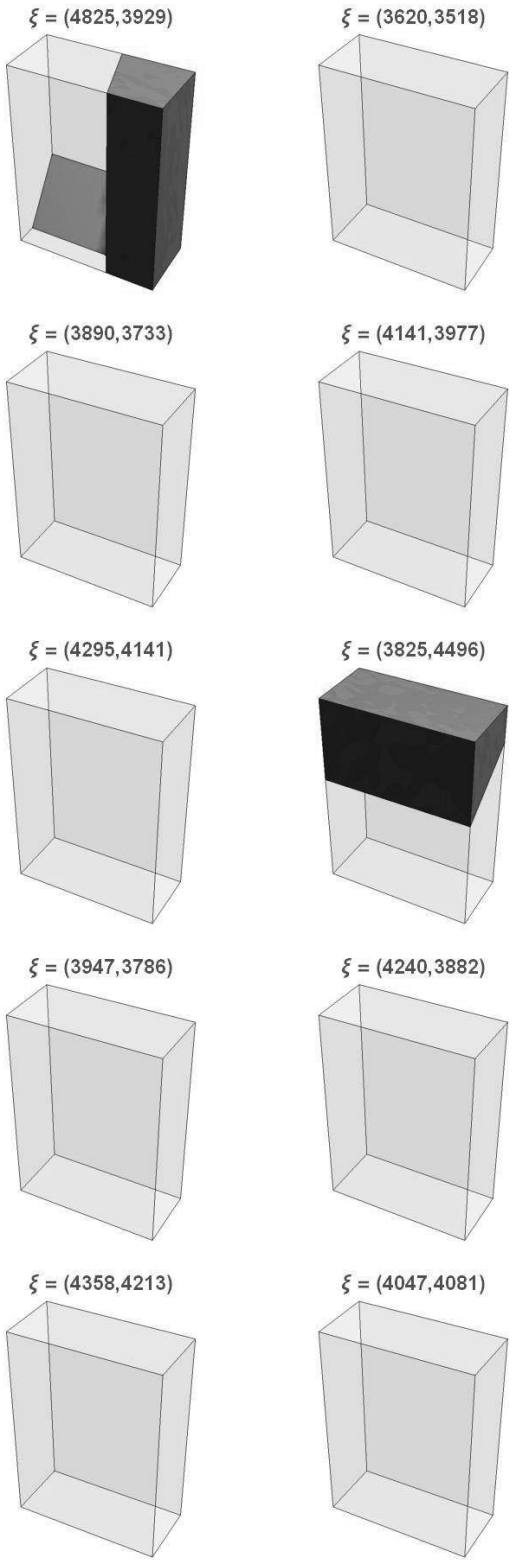


FIGURE 3.3: Simulation of ten exit load patterns along with infeasible roughness coefficients colored inside the optimal uncertainty set.

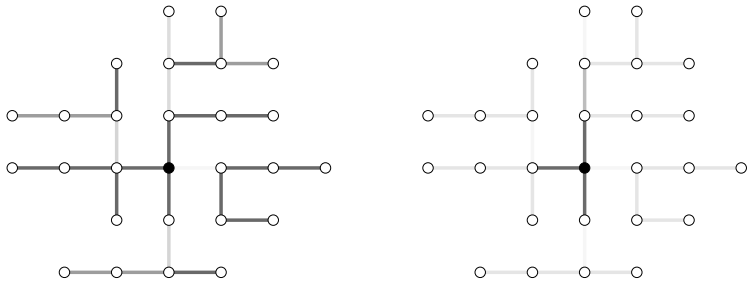


FIGURE 3.4: Computed sensitivity (displayed by color temperature scale) of the roughness coefficients for distinct pipes of a gas network with different probability levels: $\rho = 0.8$ (left) and $\rho = 0.9$ (right).

scenarios with a certain high probability ρ and offer extra capacity to new clients for which the unknown nominations are of non-stochastic nature.

This probabilistic relaxation of an originally worst-case-type requirement for feasibility gives the network owner the chance of offering a significantly larger pipeline capacity. For the given values, it may be the case that the probability of nominations being technically feasible is larger than the value ρ desired by the network owner. This degree of freedom can be used then to extend the currently booked capacities by a value that still allows one to keep the desired probability level ρ, no matter what additional nominations in the extended range have been chosen. The resulting optimization problem for the network owner was published in [4] and is presented in the following.

3.4.1 Description of the Optimization Problem

Let $G = (\mathcal{V}^+, \mathcal{E})$ be a given tree-like passive stationary gas network whose root refers to a single entry node (labeled zero) supplying the remaining nodes (exits, labeled $1, \ldots, m$) with gas. We recall that for $k, \ell \in \mathcal{V}$, denote $k \succeq \ell$ if, in G, the unique directed path from the root to k, denoted $\Pi(k)$, passes through ℓ. Moreover, the nomination vector is denoted by $\hat{b} := (b_0, b)$ with b_0 being the load at the entry and b being the exit-load. The fact that the feasibility of the original enhanced vector $\hat{b}$ is already determined by the feasibility of its exit part b in the inequality system above follows from the total load (sum of all components) always being zero. Hence, in the following we will only speak of the feasibility of the load vector at exits from the very beginning. According to [62], a vector of exit loads in this configurations is technically feasible, whenever the inequality system given in Proposition 3.1.

$$g_{k,\ell}(b) := h_k(b) + (p_k^{\max})^2 - h_\ell(b) - \left(p_\ell^{\min}\right)^2 \geq 0 \quad (k, \ell = 0, \ldots, m) \tag{3.26}$$

is satisfied. Here, $p_k^{\min}$, $p_k^{\max}$ refer to lower and upper pressure bounds in the nodes of the network and

$$h_k(b) := \begin{cases} \sum\limits_{e \in \Pi(k)} \Phi_e \left(\sum\limits_{t \succeq h(e)} b_t \right)^2 & \text{if } k \geq 1 \\ 0 & \text{if } k = 0 \end{cases} . \tag{3.27}$$

According to the model analyzed here, a concrete nomination vector of loads at the exits is given as the sum $\xi + y$ of a random vector $\xi \in [0, L]$ and a second vector $y \in [0, x]$. Here, the component L_k refers to the current upper-bound capacity at node k and x_k is an additionally booked capacity, say for a new customer at the same node k ($k = 1, \ldots, m$). The motivation of modeling ξ as a random vector comes from the fact that a sufficiently large data basis of past load nominations within the currently booked capacities may be given, which would allow one to approximate ξ as a truncated Gaussian distribution (see [81]).

While this stochastic information enables the network owner to relax the technical feasibility of exit nominations in a probabilistic sense, nothing is known in contrast about the future nomination pattern of the new customer, so that one has to be prepared principally for every possible nomination $y \in [0, x]$. This constellation leads the network owner to define a capacity extension x as feasible, whenever the constraint

$$\mathbb{P}\left(\xi \in [0, L] \,:\; g_{k,\ell}\left(\xi + y\right) \geq 0 \quad \forall y \in [0, x]\,; \quad \forall k, \ell = 0, \ldots, m\right) \geq \rho \tag{3.28}$$

is satisfied by that x. The meaning of this constraint is as follows: The capacity extension x is feasible if and only if with probability larger than $\rho \in [0, 1]$ the sum $\xi + y$ of the original random nomination vector and of a new nomination vector can be technically realized for every such new nomination vector in the limits $[0, x]$. By its structure, (3.28) is a probabilistic constraint, but it is a nonstandard one in that it contains a robust (worst-case) ingredient, which makes the given random inequality system an infinite one. Such probust constraints have been first considered in the previous section and in [60].

By regulatory law, the network operator is endorsed to maximize the capacity that can be booked. This leads him to the consideration of the following optimization problem

$$\text{maximize } w^T x \quad \text{subject to (3.28)}, \tag{3.29}$$

where w is a weighted preference vector for capacity maximization (e.g., $w =$

$(1,\ldots,1)$ in the case of no preferences among exit nodes). In the following, we will present a solution approach for problem (3.29) in the case of the random vector ξ having a truncated multivariate Gaussian distribution. This is one of the distribution instances observed in the historical data of load nominations (see [81, p. 275]).

3.4.2 Robust counterpart reformulation

In order to apply the methodology of optimization problems with probust constraints, we first have to reduce the infinite system of constraints (3.28) into a finite one. To this end we make use of the robust counterpart of the problem

$$g_{k,\ell}(\xi + y) \geq 0 \;\; \forall y \in [0, x] \quad \Longleftrightarrow \quad \min_{y\in[0,x]} g_{k,\ell}(\xi + y) \geq 0,$$

where $k, \ell = 0, \ldots, |\mathcal{V}|$ and $k \neq l$. We denote by

$$\tilde{g}_{k,\ell}(x, \xi) := \min_{y\in[0,x]} g_{k,\ell}(\xi + y), \;\; k, \ell = 0, \ldots, |\mathcal{V}|,$$

the minimum function depending on both x and ξ. For this problem, we can get an explicit representation of the minimum function, but let us first consider the inequality system

$$g_{k,\ell}(z + y) \geq 0 \quad (k, \ell = 0, \ldots, m). \tag{3.30}$$

By (3.26), we have that $g_{0,0}\,(\xi + y) = (p_0^{\max})^2 - \left(p_0^{\min}\right)^2$ so that the inequality $g_{0,0}(\xi + y) \geq 0$ always holds true automatically. As for the remaining inequalities in (3.30), we observe that, by definition,

$$\begin{aligned}
g_{k,0}\,(\xi + y) \geq 0 \quad &\Longleftrightarrow \quad \sum_{e\in\Pi(k)} \Phi_e \left(\sum_{t\succeq h(e)} \xi_t + y_t \right)^2 \geq \left(p_0^{\min}\right)^2 - (p_k^{\max})^2 \\
&\quad (k = 1, \ldots, m) \\
g_{0,\ell}\,(\xi + y) \geq 0 \quad &\Longleftrightarrow \quad \sum_{e\in\Pi(\ell)} \Phi_e \left(\sum_{t\succeq h(e)} \xi_t + y_t \right)^2 \leq (p_0^{\max})^2 - \left(p_\ell^{\min}\right)^2 \\
&\quad (\ell = 1, \ldots, m) \\
g_{k,\ell}\,(\xi + y) \geq 0 \quad &\Longleftrightarrow \quad \sum_{e\in\Pi(k)} \Phi_e \left(\sum_{t\succeq h(e)} \xi_t + y_t \right)^2 - \sum_{e\in\Pi(\ell)} \Phi_e \left(\sum_{t\succeq h(e)} \xi_t + y_t \right)^2 \\
&\quad \geq \left(p_\ell^{\min}\right)^2 - (p_k^{\max})^2 \\
&\Longleftrightarrow \quad \sum_{e\in\Pi(k)\setminus\Pi(\ell)} \Phi_e \left(\sum_{t\succeq h(e)} \xi_t + y_t \right)^2 - \sum_{e\in\Pi(\ell)\setminus\Pi(k)} \Phi_e \left(\sum_{t\succeq h(e)} \xi_t + y_t \right)^2 \\
&\quad \geq \left(p_\ell^{\min}\right)^2 - (p_k^{\max})^2 \quad (k, \ell = 1, \ldots, m)\,.
\end{aligned}$$

By the definition of $\Pi(k)$ we have that $\Pi(0) = \emptyset$. As a consequence, the inequalities above involving at least one index k or l equal to zero amount to the same inequality in the last system above (starting with $k, l = 1$) by formally allowing one or both indices being equal to zero. This observation leads to the equivalence

$$\begin{gathered} g_{k,\ell}(\xi + y) \geq 0 \quad \Longleftrightarrow \\ \sum_{e \in \Pi(k) \setminus \Pi(\ell)} \Phi_e \left(\sum_{t \succeq h(e)} \xi_t + y_t \right)^2 - \sum_{e \in \Pi(\ell) \setminus \Pi(k)} \Phi_e \left(\sum_{t \succeq h(e)} \xi_t + y_t \right)^2 \geq \left(p_\ell^{\min}\right)^2 - \left(p_k^{\max}\right)^2 \\ (k, \ell = 0, \ldots, m). \end{gathered} \tag{3.31}$$

In the next step, we want to reduce the infinite random inequality system

$$g_{k,\ell}(\xi + y) \geq 0 \quad \forall y \in [0, x]; \quad \forall k, \ell = 0, \ldots, m$$

inside (3.28) to a finite one. With $\Phi_e, \xi_t \geq 0$ for all e and t, it follows from (3.31) that, for $k, \ell = 0, \ldots, m$

$$\begin{gathered} g_{k,\ell}(\xi, y) \geq 0 \quad \forall y \in [0, x] \quad \Longleftrightarrow \\ \min_{y \in [0,x]} \left\{ \sum_{e \in \Pi(k) \setminus \Pi(\ell)} \Phi_e \left(\sum_{t \succeq h(e)} \xi_t + y_t \right)^2 - \sum_{e \in \Pi(\ell) \setminus \Pi(k)} \Phi_e \left(\sum_{t \succeq h(e)} \xi_t + y_t \right)^2 \right\} \\ \geq \left(p_\ell^{\min}\right)^2 - \left(p_k^{\max}\right)^2 \\ \Longleftrightarrow \\ \min_{y \in [0,x]} \sum_{e \in \Pi(k) \setminus \Pi(\ell)} \Phi_e \left(\sum_{t \succeq h(e)} \xi_t + y_t \right)^2 - \max_{y \in [0,x]} \sum_{e \in \Pi(\ell) \setminus \Pi(k)} \Phi_e \left(\sum_{t \succeq h(e)} \xi_t + y_t \right)^2 \\ \geq \left(p_\ell^{\min}\right)^2 - \left(p_k^{\max}\right)^2 \\ \Longleftrightarrow \\ \sum_{e \in \Pi(k) \setminus \Pi(\ell)} \Phi_e \left(\sum_{t \succeq h(e)} \xi_t \right)^2 - \sum_{e \in \Pi(\ell) \setminus \Pi(k)} \Phi_e \left(\sum_{t \succeq h(e)} \xi_t + x_t \right)^2 \geq \left(p_\ell^{\min}\right)^2 - \left(p_k^{\max}\right)^2. \end{gathered} \tag{3.32, 3.33}$$

Here, in the second-last equivalence, the separation of the overall min into a difference of min and max relies on the fact that the edges $e \in \Pi(k) \setminus \Pi(\ell)$, and $e \in \Pi(\ell) \setminus \Pi(k)$ and so the nodal sets

$$\bigcup_{e \in \Pi(k) \setminus \Pi(\ell)} \{t \succeq h(e)\}, \qquad \bigcup_{e \in \Pi(\ell) \setminus \Pi(k)} \{t \succeq h(e)\}$$

are disjoint (otherwise there would be a contradiction with the network being a tree).

Therefore, after analytically finding the robust counterpart of the inner problem, our optimization problem of maximizing booking capacities turns into a classical problem with a joint probabilistic constraint. The reformulation of (3.29) reads

$$\max w^T x \quad \text{s.t.} \quad \mathbb{P}\left(\tilde{g}_{k,\ell}(x,\xi) \geq 0,\ \forall k,\ell = 0,\ldots,|\mathcal{V}|\right) \geq \rho \tag{3.34}$$

3.4.3 Algorithmic approach to solving the capacity problem

We will suppose that the random vector ξ of stochastic exit nominations follows a truncated multivariate Gaussian distribution:

$$\xi \sim \mathcal{TN}(\mu, \Sigma, [0, L]).$$

More precisely, the distribution of ξ is obtained by truncating an m-dimensional Gaussian distribution with mean μ and covariance matrix Σ to an m-dimensional rectangle $[0, L]$ representing the (historical) booked capacity at exit node i. By definition of truncation, this means that there exists a Gaussian random vector $\tilde{\xi} \sim \mathcal{N}(\mu, \Sigma)$ such that

$$\mathbb{P}\left(\xi \in A\right) = \frac{\mathbb{P}\left(\tilde{\xi} \in A \cap [0, L]\right)}{\mathbb{P}\left(\tilde{\xi} \in [0, L]\right)}$$

holds true for all Borel measurable subsets $A \subseteq \mathbb{R}^m$. Hence, in order to determine probabilities under a truncated Gaussian distribution, it is sufficient to be able to determine probabilities under a Gaussian distribution itself. Applying this observation to the probabilistic constraint (3.28) and combining it with (3.33) yields the equivalent description

$$\mathbb{P}\Bigg(\tilde{\xi} \in [0, L] : \sum_{e \in \Pi(k) \setminus \Pi(\ell)} \Phi_e \left(\sum_{t \succeq h(e)} \xi_t\right)^2 - \sum_{e \in \Pi(\ell) \setminus \Pi(k)} \Phi_e \left(\sum_{t \succeq h(e)} \xi_t + x_t\right)^2 \geq (p_\ell^{\min})^2 - (p_k^{\max})^2;\quad \forall k, \ell = 0, \ldots, m\Bigg) \geq p \cdot \mathbb{P}\left(\tilde{\xi} \in [0, L]\right). \tag{3.35}$$

This is now, in contrast to (3.28) a conventional probabilistic constraint over a finite inequality system. In order to deal algorithmically with the probabilistic constraint (3.35), one has evidently to be able to calculate for each fixed decision vector x the probabilities occurring there, as well as their derivatives with respect to x. We briefly sketsch the methodology used here in the following.

Spheric-radial decomposition

From the well-known spheric-radial decomposition (see Theorem 2.7) of a Gaussian random vector $\tilde{\xi} \sim \mathcal{N}(\mu, \Sigma)$ it follows that the probability of an arbitrary Borel measurable subset M of $\mathbb{R}^m$ may be represented as the following integral over the unit sphere $\mathbb{S}^{m-1}$:

$$\mathbb{P}\left(\tilde{\xi} \in M\right) = \int\limits_{v \in \mathbb{S}^{m-1}} \mu_\chi\left(E(v)\right) d\mu_\eta(v).$$

Here, μ_χ refers to the one-dimensional Chi-distribution with m degrees of freedom, μ_η is the uniform distribution on $\mathbb{S}^{m-1}$ and

$$E(v) := \{r \geq 0 : \mu + rPv \in M\},$$

where P is a factor from a decomposition $\Sigma = PP^T$ of the covariance matrix Σ. Following these remarks, the probability on the left-hand side of (3.35) (depending also on the decision variable x) can be represented as

$$\int\limits_{v \in \mathbb{S}^{m-1}} \mu_\chi\left(E(v,x)\right) d\mu_\eta(v), \tag{3.36}$$

where

$$E(v,x) = \{r \geq 0 : \mu + rPv \in [0,L]\} \cap \bigcap_{k,\ell=0,\ldots,m} E^{k,\ell}(v,x) \tag{3.37}$$

and, with P_t denoting the row number t of P, for $k, \ell = 0, \ldots, m$:

$$E^{k,\ell}(v,x) := \Bigg\{ r \in \mathbb{R} \Bigg| \tag{3.38}$$

$$\sum_{e \in \Pi(k) \setminus \Pi(\ell)} \Phi_e \left(\sum_{t \succeq H(e)} \mu_t + rP_t v \right)^2 - \sum_{e \in \Pi(\ell) \setminus \Pi(k)} \Phi_e \left(\sum_{t \succeq H(e)} \mu_t + rP_t v + x_t \right)^2$$

$$\geq (p_\ell^{\min})^2 - (p_k^{\max})^2 \Bigg\}.$$

In order to evaluate the integrand in (3.36), we already know that one has to be able to characterize (for each given $v \in \mathbb{S}^{m-1}$ and $x \in \mathbb{R}^m$) the set $E(v,x)$ and to determine its Chi probability (see Chapter 2).

Hence, we are left with the task of efficiently representing $E(v,x)$ as a finite union of intervals. This is easily done for the first set in the intersection providing $E(v,x)$

in (3.37), which can be shown either to be empty or an interval:

$$\begin{aligned} &\left\{ r \geq 0 | \mu + rPv \in [0, L] \right\} = \\ &\begin{cases} \emptyset & \text{if } \exists t \in \{1, \ldots, m\} : P_t v = 0,\ \mu_t \notin [0, L] \\ [C, R] & \text{else;} \end{cases} \end{aligned} \tag{3.39}$$

$$\begin{aligned} C &:= \max \left\{ 0, \max_{P_t v > 0} \frac{-\mu_t}{P_t v}, \max_{P_t v < 0} \frac{L_t - \mu_t}{P_t v} \right\} \\ R &:= \min \left\{ \min_{P_t v > 0} \frac{L_t - \mu_t}{P_t v}, \min_{P_t v < 0} \frac{-\mu_t}{P_t v} \right\}. \end{aligned}$$

As for the second part of the intersection in (3.37), we will provide for each k, l an explicit representation of the set $E^{k,\ell}(v, x)$, either as a single interval or as the disjoint union of two intervals such that the union over all these sets (and the first set determined above) is readily obtained in the form of a finite union of disjoint intervals. Indeed, upon developing the expressions in (3.38) in terms of r, one arrives at the representation

$$E^{k,\ell}(v, x) = \{ r \in \mathbb{R} | \alpha^{k,\ell} r^2 + \beta^{k,\ell} r + \gamma^{k,\ell} \geq 0 \} \quad (k, \ell = 0, \ldots, m),$$

where, for $k, \ell = 0, \ldots, m$:

$$\begin{aligned} \alpha^{k,\ell} &:= \sum_{e \in \Pi(k) \setminus \Pi(\ell)} \Phi_e \left(\sum_{t \succeq H(e)} P_t v \right)^2 - \sum_{e \in \Pi(\ell) \setminus \Pi(k)} \Phi_e \left(\sum_{t \succeq H(e)} P_t v \right)^2 \\ \beta^{k,\ell} &:= 2 \sum_{e \in \Pi(k) \setminus \Pi(\ell)} \Phi_e \left(\sum_{t \succeq H(e)} \mu_t \right) \left(\sum_{t \succeq H(e)} P_t v \right) \\ &\quad - 2 \sum_{e \in \Pi(\ell) \setminus \Pi(k)} \Phi_e \left(\sum_{t \succeq H(e)} \mu_t + x_t \right) \left(\sum_{t \succeq H(e)} P_t v \right) \\ \gamma^{k,\ell} &:= \sum_{e \in \Pi(k) \setminus \Pi(\ell)} \Phi_e \left(\sum_{t \succeq H(e)} \mu_t \right)^2 - \sum_{e \in \Pi(\ell) \setminus \Pi(k)} \Phi_e \left(\sum_{t \succeq H(e)} \mu_t + x_t \right)^2 \\ &\quad + (p_k^{\max})^2 - (p_\ell^{\min})^2. \end{aligned}$$

This leads, by case distinction and elementary calculus, to the following explicit representation of $E^{k,\ell}(v,x)$ for $k,\ell = 0,\dots,m$:

$$E^{k,\ell}(v,x) =$$

$$\begin{cases} \emptyset & 1) \text{ or } 2) \\ \mathbb{R} & 3) \text{ or } 4) \\ \left[-\frac{\gamma^{k,\ell}}{\beta^{k,\ell}},\infty\right) & 5) \\ \left(-\infty,-\frac{\gamma^{k,\ell}}{\beta^{k,\ell}}\right] & 6) \\ \left(-\infty,\frac{-\beta^{k,\ell}-\sqrt{\left(\beta^{k,\ell}\right)^2-4\alpha^{k,\ell}\gamma^{k,\ell}}}{2\alpha^{k,\ell}}\right] \cup \left[\frac{-\beta^{k,\ell}+\sqrt{\left(\beta^{k,\ell}\right)^2-4\alpha^{k,\ell}\gamma^{k,\ell}}}{2\alpha^{k,\ell}},\infty\right) & 7) \\ \left[\frac{-\beta^{k,\ell}+\sqrt{\left(\beta^{k,\ell}\right)^2-4\alpha^{k,\ell}\gamma^{k,\ell}}}{2\alpha^{k,\ell}},\frac{-\beta^{k,\ell}-\sqrt{\left(\beta^{k,\ell}\right)^2-4\alpha^{k,\ell}\gamma^{k,\ell}}}{2\alpha^{k,\ell}}\right] & 8) \end{cases},$$

where the case study is done according to

$$\begin{aligned} 1) &\quad \alpha^{k,\ell} = \beta^{k,\ell} = 0,\ \gamma^{k,\ell} < 0 \\ 2) &\quad \alpha^{k,\ell} < 0,\ \left(\beta^{k,\ell}\right)^2 < 4\alpha^{k,\ell}\gamma^{k,\ell} \\ 3) &\quad \alpha^{k,\ell} = \beta^{k,\ell} = 0,\ \gamma^{k,\ell} \geq 0 \\ 4) &\quad \alpha^{k,\ell} > 0,\ \left(\beta^{k,\ell}\right)^2 < 4\alpha^{k,\ell}\gamma^{k,\ell} \\ 5) &\quad \alpha^{k,\ell} = 0,\ \beta^{k,\ell} > 0 \\ 6) &\quad \alpha^{k,\ell} = 0,\ \beta^{k,\ell} < 0 \\ 7) &\quad \alpha^{k,\ell} > 0,\ \left(\beta^{k,\ell}\right)^2 \geq 4\alpha^{k,\ell}\gamma^{k,\ell} \\ 8) &\quad \alpha^{k,\ell} < 0,\ \left(\beta^{k,\ell}\right)^2 \geq 4\alpha^{k,\ell}\gamma^{k,\ell}. \end{aligned}$$

Along with (3.39) we may use this explicit description in order to efficiently represent the set $E(v,x)$ in (3.37) as the desired finite union of intervals.

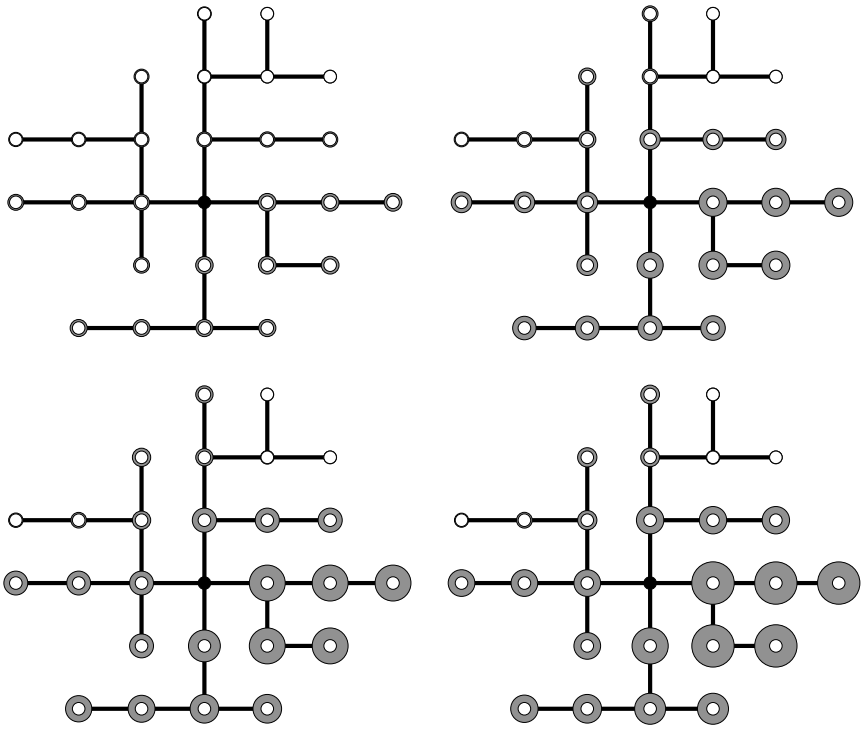
It is important to note that, at the same time, the partial derivatives of the probability with respect to the decision variable x can be calculated as a spherical integral of type (3.36) again, however with a different integrand, which is easily obtained from the partial derivatives of the initial data [141]. In this gradient formula, the same disjoint union of intervals as in the computation of the probability itself is employed. The spherical integrals can be approximated by finite sums using quasi-Monte Carlo sampling on the sphere (see, e.g., [25] or Chapter 2). Then, for each

sampled direction v on the sphere, one may update first the probability itself and then, simultaneously, the gradient of the probability with respect to x by using the same disjoint union of intervals in both cases. This approach makes the gradient come almost for free as far as computation time is concerned. Having access to values and gradients of the probabilistic constraint (3.35), one may set up an appropriate nonlinear optimization solver for solving (3.28). For the subsequent numerical results, we employed a simple projected gradient method.

3.4.4 Numerical results for an example

As an illustrating example we considered a network as displayed in Fig. 3.5 with one entry (filled black circle) and 26 exits. The parameters of the network (i.e., pressure bounds, roughness coefficients, truncated Gaussian distribution for the random nominations at exits) were chosen in realistic quantities that are modified versions of real data.

FIGURE 3.5: Solution of the capacity maximization problem at exits for different probability levels: 0.95 (top left); 0.9 (top right); 0.85 (bottom left); 0.8 (bottom right).



We didn't assume any preferences in the allocation of new capacities, so the weight vector in the objective of (3.28) was chosen as $w := \mathbf{1}$. The colored rings

around exit points refer to the optimal cumulative capacities (historical+new), i.e., $L + x$ after maximization, upon choosing probability levels $p = 0.95, 0.9, 0.85, 0.8$. It can be clearly seen how decreasing of the probability level allows for increasing the allocation of apacity in certain regions of the network.

FIGURE 3.6: Two scenarios for random exit loads ξ according to the chosen multivariate truncated Gaussian distribution. Left: feasible scenario; Right: infeasible scenario.

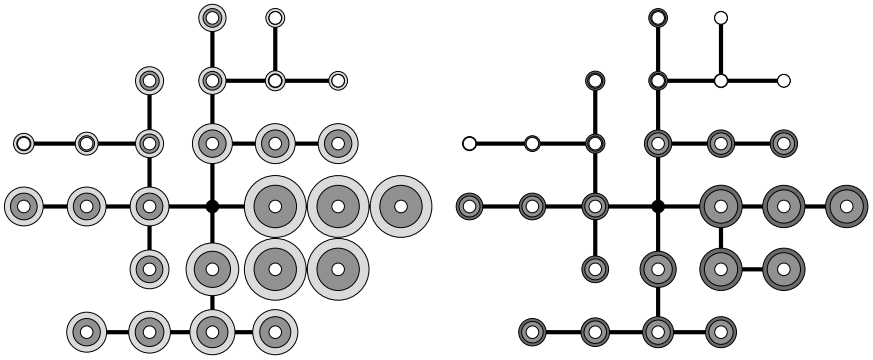


Fig.3.6 illustrates how the computed solution for a probability level $p = 0.8$ works for two random exit nomination scenarios ξ simulated a posteriori according to the chosen truncated multivariate Gaussian distribution. The first scenario is feasible because one could uniformly add a common capacity to every exit (green color) in order to satisfy this scenario. In contrast, the second scenario is infeasible because one would have to (uniformly) reduce the capacities by an amount corresponding to the dark red rings in order to satisfy this scenario. When simulating a large set of such scenarios, say 1000, it would turn out that according to the probability level $p = 0.8$ approximately 800 are feasible, while 200 are infeasible.

In the presence of a network with cycles, the methodology presented here does not work. In this case, one cannot analytically identify the worst-case solution of the robust subproblem. In general, for an optimization problem under probust functions, one has to solve a semi-indefinite subproblem. Moreover, if the uncertainty set depends on the decision policy (it can be seen as a moving index set, in contrast with the multiple probabilistic function with fixed set studied in [141]), then the spherical-radial decomposition algorithm and sub-(gradients) have to be applied for Banach spaces (see [143]).

Chapter 4

Dynamic Joint Probabilistic Constraints

In this chapter, we present our investigation on the structural properties of the *dynamic joint probabilistic constraints* in two different Banach spaces. For a two-stage linear and bilinear model, we verify Lipschitz continuity and differentiability of the probability functions. Moreover, we provide explicit derivative formulae. We further prove the existence of a solution for a hydro-power reservoir optimization problem in the $W^{1,2}$ space. Finally, numerical results for the solution of the two-stage problem for a hydro-power reservoir in L^2 space are illustrated.

To solve an optimization problem with dynamic joint probabilistic constraints, we transform it into an equivalent deterministic optimization problem in a Banach space. That is why, in the next section, we start with a brief overview of deterministic nonlinear constrained optimization in an infinite-dimensional context.

4.1 Constrained Optimization in Banach Spaces

The aim of this section is to provide a preliminary overview of the tools and methods of infinite-dimensional optimization, which will be used in the rest of the chapter. For the interested reader we recommend [21, 29]. Moreover, [85] discusses some fundamental differences and common properties between infinite and finite-dimensional programming.

We can write an optimization problem in an abstract space $\mathcal{X}$ as

$$\min_{x \in \mathcal{C}} J(x)$$

where $\mathcal{C} \subset \mathcal{X}$ and $J : \mathcal{C} \to \bar{\mathbb{R}}$ is the objective functional.

4.1.1 Existence of Optima

The first and most fundamental question in optimization theory is whether the maximum (or minimum) of a function exists on a given set. We know from finite-dimensional spaces that two basic properties are needed: compactness and lower semicontinuity. However, compactness with respect to the norm topology is too restrictive in an infinite dimensional Banach space. Instead, [29, Thm. 3.17] suggests that we should work on a reflexive Banach space and with weak convergence, to prove the existence of a solution in an optimization problem. Before stating the existence theorem for optimization problems in infinite dimensional spaces, we will need the following notions:

Definition 4.1. Let A be a subset of a normed space X. A minimizing sequence for the optimization problem $\min_{x\in A} J(x)$ is a sequence $(x_n)_{n\in\mathbb{N}} \subset A$ satisfying

$$\lim_{n\to+\infty} J(x_n) = \inf_{x\in A} J(x).$$

Definition 4.2. Let $(\mathcal{X}, ||\cdot||_{\mathcal{X}})$ be a reflexive Banach space and $A \subset \mathcal{X}$ a subspace. Then the functional $J : A \to \mathbb{R}\cup\{+\infty\}$ is *coercive* provided that $J(x) \to +\infty$ if $||x||_{\mathcal{X}} \to \infty$.

Definition 4.3. $J : \mathcal{X} \to \bar{\mathbb{R}}$ is said to be *sequentially weakly lower semicontinuous* over $\mathcal{X}$ provided that for every sequence $(x_n)_{n\in\mathbb{N}} \subset \mathcal{X}$ that converges weakly $x_n \rightharpoonup \bar{x}$,

$$J(\bar{x}) \leq \liminf_{n\to\infty} J(x_n).$$

We now have all concepts in hand to prove the central existence result in infinite dimensional Banach spaces. This is the *direct method* of calculus of variations [98, p.32], invented originally in [130]. In the following we prove it with a more relaxed hypothesis (sequentially weak properties). We will use this proposition in the last section of this chapter.

Proposition 4.4. *Let $\mathcal{X}$ be a reflexive Banach space and $A \subset \mathcal{X}$ be a nonempty sequentially weakly closed subset of $\mathcal{X}$. Assuming*

i) $J : A \to \mathbb{R}\cup\{+\infty\}$ is proper, coercive,

ii) J is sequentially weakly lower semi-continuous,

J is bounded from below on A and attains its infimum in A.

Proof. Consider a minimizing sequence $(x_n)_{n\in\mathbb{N}} \subset A$ with

$$\lim_{n\to\infty} J(x_n) = \inf_{x\in A} J(x) \geq -\infty.$$

Since J is proper, we have that $\inf_{x\in A} J(x) < \infty$. Then, $J(x_n)$ is bounded from below and by coercitivity (since $\lim_{n\to\infty} J(x_n) \neq +\infty$ then $\lim_{n\to\infty} ||x_n|| \neq +\infty$) we have that $(x_n)_{n\in\mathbb{N}}$ is a bounded sequence. Since $\mathcal{X}$ is a reflexive Banach space, there exists a weakly convergent subsequence $(x_{n_k})_{k\in\mathbb{N}}$ with $x_n \rightharpoonup \bar{x}$ (from [29, Thm. 3.17]). Because A is sequentially weakly closed, we have that $\bar{x} \in A$. From the definition of the minimizing sequence together with the sequentially weakly lower semicontinuity of J and the definition of infimum we obtain

$$\inf_{x\in A} J(x) \leq J(\bar{x}) \leq \liminf_{k\to\infty} J(x_{n_k}) = \inf_{x\in A} J(x) < +\infty.$$

This implies that the infimum is attained in $\bar{x}$, which is therefore the desired minimizer. □

4.1.2 KKT in Banach spaces

Consider the following generic nonlinear optimization problem:

$$\min_{x\in Q} \quad J(z) \quad \text{s.t.} \quad G(z) \in K, \tag{4.1}$$

where $J : Z \to \mathbb{R}, G : Z \to Y$ and $Q \subseteq Z$ and $K \subseteq Y$ are nonempty closed convex subsets. We assume that Z and Y are Banach spaces and that $J(z), G(z)$ are continuously differentiable (and hence Frechét differentiable). We say that a point $z \in Z$ is *feasible* if $z \in Q$ and $G(z) \in K$ and denote by

$$\mathcal{M} := Q \cap G^{-1}(K) = \{z \in Q : G(z) \in K\} \tag{4.2}$$

the feasible set of (4.1). Note that K is assumed to be a convex set, but we make no convexity assumptions on the mapping G. In particular, the feasible set $\mathcal{M}$ may not be convex.

We define the Lagrange function of (4.1), which is the function

$$\mathcal{L} : Z \times Y^* \to \mathbb{R}, \quad \mathcal{L}(z, \lambda) := J(z) + \langle \lambda, G(z) \rangle.$$

Next, we define some basic objects, as in Section 1.1, which are useful when characterizing the geometric structure of optimization problems in Banach spaces.

Definition 4.5. Let $D \subset Z$ be a nonempty set. We say that D is a cone if $\alpha D \subset D$ for all $\alpha > 0$.

Definition 4.6. Let $D \subset Z$ be an arbitrary set. Then the *polar cone* of D, denoted by D°, is described by

$$D^\circ := \{z \in Z^* : \langle z, d\rangle \leq 0 \ \forall d \in D\}.$$

Definition 4.7. Let $C \subseteq Z$ be a convex set. We define the *normal cone* as

$$N_C(z) := \{w \in Z^* : \langle w, y - z\rangle \leq 0 \ \forall y \in D\}.$$

Definition 4.8. A pair $(x^*, \lambda^*) \in X \times Y^*$ is a KKT point of (4.1) if

$$-\nabla \mathcal{L}(x^*, \lambda^*) \in N_Q(x^*), \ \textit{and} \ \ \lambda^* \in N_K(G(x^*)).$$

We say that $x^* \in X$ is a stationary point of (4.1) if (x^*, λ^*) is a KKT point for some multiplier $\lambda^* \in Y^*$, and denote by $\Lambda(x^*)$ the set of such multipliers.

Note that the above implies that $x^* \in Q$ and $G(x^*) \in K$; otherwise at least one of the corresponding normal cones would be empty. Thus, every stationary point of (4.1) is necessarily feasible.

Remark 4.9. Assume that $K \subseteq Y$ is a closed convex cone. Then the inclusion $\lambda^* \in \mathcal{N}_K(G(x^*))$ in the KKT conditions can equivalently be stated as

$$G(x^*) \in K, \ \lambda^* \in K^\circ, \ \text{and} \ \ \langle \lambda^*, G(x^*)\rangle = 0.$$

These three conditions are called *complementary conditions*. We can interpret these conditions as the finite dimensional complementary conditions: $G(x^*) \geq 0, \lambda^* \geq 0$ (in the dual sense), and their product being equal to zero.

The same way as in finite dimension, we also need certain regularity properties for the KKT conditions to be necessary for (4.1). As we learned in Theorem 1.9, such properties are called *constrained qualifications*. The constraint qualification we will use in the setting of infinite-dimensional spaces is the following.

Definition 4.10. Let $x \in X$ be a feasible point for (4.1). We say that the *Robinson constraint qualification* (RCQ) holds in x if G is continuously Fréchet differentiable at point x and if

$$0 \in \text{int}\{G(x) + \nabla G(x)(Q - x) - K\}.$$

The above condition was introduced by [114]. A more detailed study of RCQ, its consequences and its characterizations for optimization problems with perturbations, can be read in [21, ch.3].

Theorem 4.11. *(KKT conditions under RCQ, [21, Thm. 3.9]) Let x^* be a local minimizer of* (4.1) *and assume that RCQ holds in x^*. Then the set of Lagrange multipliers $\Lambda(x^*)$ is non-empty, closed, convex and bounded in Y^*.*

The following result is precisely the geometric property that lies at the core of the KKT conditions.

Corollary 4.12. *[21, Cor. 2.91] Suppose that a mapping $G : X \to Y$ is continuously differentiable at a feasible point $x_0 \in \mathcal{M}$ and assume that RCQ holds in x_0. Then*

$$T_{\mathcal{M}}(x_0) = \{d \in T_Q(x_0) : \langle \nabla G(x_0), d \rangle \in T_K(G(x_0))\}.$$

Finally, we note that in the case of a nonlinear optimization problem in finite dimensions, when $Y = \mathbb{R}^n \times \mathbb{R}^k$ and $K = \mathbb{R}^n_-$, Robinson's regularity condition reduces to the Mangasarian-Fromovitz constraint qualification [88] and the characterization of the tangent cone reduces to (1.4).

In the infinite-dimensional case, for the first order necessary conditions for the problem with differentiable maps see [21, 66]. In [28] one can find the KKT theorem for the problem with a finite number of nonlinear inequality constraints in a Banach space under the linear independence constraint qualification. Second-order conditions in infinite-dimensional spaces are rather recent. For the differentiable case see [76].

4.2 Dynamic Joint Probabilistic Constraints

In chapter 1.3.3 we saw that the prototype for an optimization problem with joint probabilistic constraints has the form

$$\begin{array}{ll} \max\limits_{x} & f(x) \\ \text{s.t.} & \mathbb{P}\left(g(x,\xi) \geq 0\right) \quad \leq \quad p, \end{array} \tag{4.3}$$

where $x \in \mathbb{R}^n$ is a decision vector, $f : \mathbb{R}^n \to \mathbb{R}$, $\xi : \Omega \to \mathbb{R}^d$ is an d-dimensional random vector defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$, $g : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^k$ is a vector-valued mapping and $p \in [0,1]$ is some probability level chosen exogenously. Accordingly, a decision vector x is feasible if $h(x,\xi) \geq 0$ is satisfied at least with probability p.

In the setting presented above, the decision is determined before the realization of the random parameter is observed (*here-and-now* decisions) and the decisions are made once and forever, making (4.3) a *static* optimization model. However, many real-world problems have parameters in the objective function and in the constraints, which may change over time, and we are interested in taking into account these changes in the optimization process. That is, we would like to take a *dynamic* decision, i.e., a process that evolves in time, for which an increasing amount of information is collected, due to the successive realization of a family of random variables.

We focus on the type of optimization problems with dynamic objective function and dynamic probability constraints. The general form of this problem is as follows:

$$\begin{aligned} &\max_{\chi} J(\chi) \\ \text{s.t.} \quad &\mathbb{P}\left(h(x_1, \chi_2(\xi_1), \chi_3(\xi_1, \xi_2) \ldots, \chi_T(\xi_1, \ldots, \xi_{T-1}), \xi_1, \ldots, \xi_T) \leq 0\right) \quad \geq \quad p, \end{aligned} \tag{4.4}$$

where $\xi := (\xi_1, \ldots, \xi_T)$ is an T-dimensional discrete time stochastic process on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Moreover,

$$\chi := (x_1, \chi_2(\xi_1), \chi_3(\xi_1, \xi_2), \ldots, \chi_T(\xi_1, \ldots, \xi_{T-1}))$$

is an adapted process of decision policies depending on previously observed outcomes of the random process. We observe that $x_1 \in \mathbb{R}$ and each component $\chi_i : \mathcal{S}_{i-1} \to \mathbb{R}$, $i = 2, \ldots, T$ is an $\mathbb{R}$-valued function defined in

$$\mathcal{S}_i := int\left(\text{Supp}(\xi_1, \ldots, \xi_i)\right) = \{x \in \mathbb{R}^i : g_{\xi_{[i]}}(x) > 0\}.$$

Our objective function $J : \mathcal{X} \to \mathbb{R}$, is defined in some Banach space of functions $\mathcal{X}$. Lastly, $h : \mathbb{R}^T \times \mathbb{R}^T \to \mathbb{R}^s$ is the constrained map.

Other investigations have tackled dynamic stochastic optimization problems through multi-stage stochastic programming, where decisions are also taken prior to the realization of the random variables, then recourse or corrective measures are taken to compensate for possible violation of constraints. For instance, a power generating company may optimize production and recourse for buying energy on the liberalized market when faced with unforeseen increase of electrical load. In contrast, a dynamic joint probabilistic constraint model can be applied when monetary compensations for violations of random realizations do not exist. For example, an isolated power generating company, so-called *mini-grid utility*, which are popular

systems to electrify developing countries [18], cannot buy or sell extra energy when confronted with unexpected increase/decrease of demand.

As mentioned in section 1.3.2, multi-stage stochastic programming is solved in a different setting, since ξ is discretized. Besides, in contrast to (1.14), the dynamic probabilistic constraint scheme makes a closed-loop decision policy ending with randomness

$$\boxed{x_1 \curvearrowright \xi_1 \curvearrowright x_2(\xi_1) \curvearrowright \xi_2 \cdots \curvearrowright x_T(\xi_1, \ldots, \xi_{T-1}) \curvearrowright \xi_T.}$$

Other authors solve stochastic optimization problems making use of stochastic optimal control with an assumption of independent components and imposing simplifications of a discrete distribution (scenario) see, e.g., [30]. For a detailed account of the theory of stochastic optimal control, we refer the reader to the classical monographs [148, 50]. Additionally, it is also worth mentioning that the authors in [49] examine the structural properties of joint probabilistic constraints in infinite-dimensional Banach spaces, but only for the *static* case.

To our knowledge, dynamic joint probabilistic constraints have been first investigated in [5], where the decision policies $x(\xi)$ are approximated by a piecewise constant function, and it turns out that the optimal policies of a reservoir problem are not linear. Also, in [65], in order to transform the dynamic joint probabilistic constraint optimization problem into a numerically tractable problem in finite dimensions, the decision policies are parametrized to linear decision rules for the underlying Gaussian or truncated Gaussian distributions. In this way, the approximating problems are analytically computed.

In this investigation, we first study the structural properties of the dynamic joint probabilistic constraints in two different Banach spaces. For a two-stage linear and bilinear model, we verify Lipschitz continuity and differentiability of the probability functions. Moreover, we provide explicit derivative formulae. We further prove the existence of a solution for a hydro-power reservoir optimization problem in the $W^{1,2}$ space. Finally, numerical results for the solution of two-stage problem for a hydro-power reservoir in the L^2 space are illustrated.

4.3 Motivating Example: Hydro-Power Operation Model

The motivation for our study of (4.4) stems from a multi-stage decision management optimization problem on a single water reservoir for hydroelectricity generation with uncertain water inflow.

Hydroelectricity is a dispatchable, flexible, scarce energy. Electricity production is said to be dispatchable when it can be controlled (increased, decreased, switched on or off). Control of water release for dam hydroelectricity allows for the control of turbine spinning and electricity generation. Also, operations of hydroelectric plants are quite flexible: they can be launched or shut down very quickly, as technical specifications allow relatively fast adaptations. However, water is a scarce resource. The water levels in reservoirs depend on the inflows coming from seasonal precipitations. For a detailed study on the economics of hydro-power we recommend [51] and for the engineering and operation [146].

The problem of the operator is to decide on an optimal release policy of water, considering technical, economic and environmental aspects. By $\xi := (\xi_1, \dots, \xi_T)$, we denote a discrete scalar random vector indicating the stochastic water inflow to the reservoirs at time periods $1, 2, \dots, T$. The role of the dams is to allow storage and transfer of electricity for future periods. The key economic question in hydro-power production is: "how much water should I release today, or store to use for next period?" The operation of hydro-power is thus a dynamic one. Taking into account the information, due to the successive realization of random inflows, we denote our release policy by

$$x := (x_1, x_2(\xi_1), x_3(\xi_1, \xi_2), \dots, x_T(\xi_1, \dots, \xi_{T-1})).$$

On the technical side, assuming no waste of water in the production of electricity, the reservoir dynamics is given by

$$l_t = l_{t-1} + \xi_t - x_t(\xi_1, \dots, \xi_{t-1}),$$

where l_t is the active reservoir storage at the end of period t, and we take $t = 0, 1, \dots T$. This continuity equation implies that

$$\begin{aligned} l_1 =& l_0 + \xi_1 - x_1 \\ l_2 =& l_1 + \xi_2 - x_2(\xi_1) \\ =& l_0 + \xi_1 + \xi_2 - x_1 - x_2(\xi_1) \\ & \vdots \\ l_t =& l_0 + \sum_{i=1}^{t} \xi_i - \sum_{i=1}^{t} x_i(\xi_1, \dots, \xi_{i-1}). \end{aligned} \tag{4.5}$$

Changing water levels may create problems for the fauna and agriculture. The

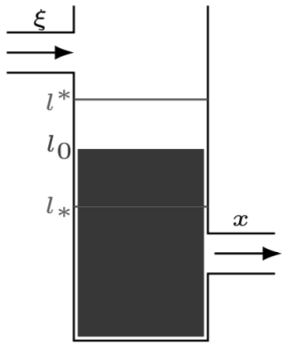


FIGURE 4.1: Illustration of a simple hydro-reservoir with random inflow ξ and decision policy χ.

reservoir has also to operate in order to mitigate environmental hazards. Thus, given an initial water level l_0 in the active reservoir, we require that during all periods of time, the water level remains between an upper value l^* (flood reserve) and some lower value l_* (dead storage). That is,

$$l_* \leq l_t \leq l^*.$$

Substituting from (4.5),

$$l_* \leq l_0 + \sum_{i=1}^{T} \xi_i - \sum_{i=1}^{T} \chi_i(\xi_1, \ldots, \xi_{i-1}) \leq l^*. \tag{4.6}$$

We do not allow for negative releases, that is $\chi_t(\xi_1, \ldots, \xi_{t-1}) \geq 0$ for all periods $t = 1, \ldots, T$.

Our model is now represented by

$$\begin{aligned} &\max_{\chi \in \mathcal{X}} J[x_1, ..., \chi_T] \\ \text{s.t.} \quad &\mathbb{P}\left(l_* \leq l_t\left(x_1, ..., \chi_t(\xi_1, ..., \xi_{t-1}), \xi_1, ..., \xi_t\right) \leq l^*,\ \forall t = 1, ..., T\right) \geq p \\ &x_1, \chi_2(\xi_1), \ldots, \chi_T(\xi_1, ..., \xi_{T-1}) \geq 0 \quad \mathbb{P}\text{-almost surely,} \end{aligned} \tag{4.7}$$

where $\mathcal{X}$ is some Banach space of all $\chi = (x_1, ..., \chi_T)$, our decision policies, and $J : \mathcal{X} \to \mathbb{R}$ a convenient objective function. For a review on different optimization models to manage hydro-reservoir see [52].

4.4 Structural Properties of Dynamical Probabilistic Constraints

In this section, we compare and contrast some topological properties of the dynamic joint probabilistic constraints in two different convenient spaces of functions.

4.4.1 Space of policy $\mathcal{X}$ and $\mathcal{X}^1$

We introduce the function spaces $\mathcal{X}$ and $\mathcal{X}^1$ that will be suitable for the subsequent developments in this paper. The function spaces to be discussed include Lebesgue and Sobolev spaces, associated with a closed bounded domain $I \subset \mathbb{R}$.

Consider $\xi_i : \Omega \to \mathbb{R}$ random variables for $i = 1, ..., T$. We assume that $\xi_{[i]} = (\xi_1, ..., \xi_i)$ has density with respect to the Lebesgue measure, denoted by $g_{\xi_{[i]}}$. We denote by $\mathcal{S}_{[i]}$ the support of the random vectors $\xi_{[i]}$, that is to say, $\mathcal{S}_{[i]} = \{x \in \mathbb{R}^i : g_{\xi_{[i]}}(x) > 0\}$. The function spaces to work with are

$$\mathcal{X} = \mathbb{R} \times L^p(\mathcal{S}_1) \times \cdots \times L^p(\mathcal{S}_{T-1})$$

and

$$\mathcal{X}^1 = \mathbb{R} \times W^{p,1}(\mathcal{S}_1) \times \cdots \times W^{p,1}(\mathcal{S}_{T-1}).$$

In the spaces $\mathcal{X}$ and $\mathcal{X}^1$ we can chose any equivalent norm in the product space. Nevertheless, for simplicity we consider the *max*-norm in $\mathcal{X}$ and $\mathcal{X}^1$ respectively, that is,

$$\|x\|_{\mathcal{X}} = \max_{i=1,...,T} \|x_i\|_r,$$
$$\|x\|_{\mathcal{X}^1} = \max_{i=1,...,T} \|x_i\|_{1,r},$$

where $\|x_i\|_r$ and $\|x_i\|_{1,r}$, denote the usual norms of x_i in $L^r(\mathcal{S}_i)$ and $W^{r,1}(\mathcal{S}_1)$, respectively. It is well-known that $L^r(\mathcal{S}_1)$ and $W^{r,1}(\mathcal{S}_1)$ are Banach spaces (see, e.g., [3]). Moreover, since $\mathcal{X}$ and $\mathcal{X}^1$ are product of Banach spaces they are also Banach spaces (see, e.g., [29]).

It is important to mention that the space of policy $\mathcal{X}$ could be defined in arbitrary sets $\mathcal{S}_{[i]}$. Nevertheless, in order to define the space of policy $\mathcal{X}^1$, we need to assume (at least) that the set $\mathcal{S}_{[i]}$ is open.

4.4.2 Continuity of φ

We provide general criteria for the continuity of the dynamic probabilistic constraint

$$\varphi(x_1, ..., x_n) = \mathbb{P}(h_i(x_1, x_2(\xi_1), ..., x_T(\xi_{[T-1]}), \xi_{[T]}) \leq 0, \quad (i = 1, \dots, k)), \tag{4.8}$$

in the space $\mathcal{X}$ and $\mathcal{X}^1$, where $h_i : \mathbb{R}^T \times \mathbb{R}^T \to \mathbb{R}$ is are given functions. It is straightforward to show that if each h_i is measurable, then (4.8) is well-defined. In order to simplify the notation let us denote $h : \mathbb{R}^T \times \mathbb{R}^T \to \mathbb{R}$ as

$$h = \max_{i=1,...,k} h_i.$$

Hence, our probability function given in (4.8) can be expressed

$$\varphi(x_1, ..., x_n) = \mathbb{P}(h(x_1, x_2(\xi_1), ..., x_T(\xi_{[T-1]}), \xi_{[T]}) \leq 0), \tag{4.9}$$

Let us start by recalling the generalization of Prekopa's theorem for the static joint probabilistic function in the infinite-dimensional setting (see, e.g., [49, Proposition 4]) and try to translate the desired results into the dynamic chance constraint setting (4.9). For this reason, we take the simple dynamic chance constraint given by $\mathbb{P}(x(\xi) \leq 1)$, with ξ a 1-dimensional random vector. Taking into account [49, Proposition 4]), in order to deal with this probability function, we can naturally try to define the function $h : L^2[a, b] \times \mathbb{R} \to \mathbb{R}$, $h(x, z) = x(z)$. However, this function is not well-defined, since the function x is defined only almost everywhere in an equivalent class. Secondly, we would have to prove that $h(x, z) = x(z)$ is quasiconcave in $L^2[a, b] \times \mathbb{R}$. Lastly, even if h is linear (or quasiconcave) we will give an example in Lemma 4.1 where the feasible set is not a convex set in the space $\mathcal{X}$. Nevertheless, we can provide different criteria for continuity of φ in the spaces $\mathcal{X}$ and $\mathcal{X}^1$.

We first provide the following two technical results.

Proposition 4.13. *Assume that ξ has a density. Consider φ defined in* (4.9) *and assume that h is lower-semicontinous. Consider a sequence $(x_1^{(n)}, x_2^{(n)}, ..., x_T^{(n)})$ in $\mathcal{X}$ (respectively in $\mathcal{X}^1$), which converges almost everywhere to $(x_1, x_2, ..., x_T)$. Then*

$$\limsup_{n\to\infty} \varphi(x_1^{(n)}, x_2^{(n)}, \dots, x_T^{(n)}) \leq \varphi(x_1, ..., x_T).$$

Moreover, if h is continuous and for all $u \in \mathbb{R}^{2T-1}$ the set $\{z \in \mathbb{R} : h(u,z) = 0\}$ has measure zero, then

$$\lim_{n\to\infty} \varphi(x_1^{(n)}, x_2^{(n)}, ..., x_T^{(n)}) = \varphi(x_1, ..., x_T). \tag{4.10}$$

Proof. Consider the sets

$$\begin{aligned} A_n :=& \{\omega \in \Omega : h\left(x_1^{(n)}, x_2^{(n)}(\xi_{[1]}(\omega)), ..., x_T^{(n)}(\xi_{[T-1]}(\omega)), \xi_{[T]}(\omega)\right) \leq 0\}, \\ A :=& \{\omega \in \Omega : h\left(x_1, x_2(\xi_{[1]}(\omega)), ..., x_T(\xi_{[T-1]}(\omega)), \xi_{[T]}(\omega)\right) \leq 0\}, \end{aligned}$$

together with the functions $\mathbb{1}_{A_n}$ and $\mathbb{1}_A$. Then, by Fatou's Lemma (see, e.g., [20]),

$$\begin{aligned} \mathbb{E}(\liminf \mathbb{1}_{A_n}) &\leq \liminf \mathbb{E}(\mathbb{1}_{A_n}) = \liminf \mathbb{P}(A_n) \\ &= \liminf \varphi(x_1^{(n)}, x_2^{(n)}, ..., x_T^{(n)}) \leq \limsup \varphi(x_1^{(n)}, x_2^{(n)}, ..., x_T^{(n)}) \\ &= \limsup \mathbb{P}(A_n) = \limsup \mathbb{E}(\mathbb{1}_{A_n}) \\ &\leq \mathbb{E}(\limsup \mathbb{1}_{A_n}). \end{aligned} \tag{4.11}$$

Now, let us show that $\limsup \mathbb{1}_{A_n} \leq \mathbb{1}_A$ $\mathbb{P}$-a.s. If the latter is true, then together with (4.11) it would follow that

$$\lim_{n\to\infty} \varphi(x_1^{(n)}, ..., x_T^{(n)}) = \limsup \mathbb{P}(A_n) \leq \mathbb{E}(\limsup \mathbb{1}_{A_n}) \leq \mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A) = \varphi(x_1, ..., x_T). \tag{4.12}$$

First, since the sequence $(x_1^{(n)}, x_2^{(n)}, ..., x_T^{(n)})$ converges almost everywhere to $(x_1, x_2, ..., x_T)$ and since $\xi_{[T]}$ has density with respect to the Lebesgue measure, we can assume w.l.o.g. that

$$(x_1^{(n)}, x_2(\xi_1(\omega)), ..., x_T^{(n)}(\xi_{[T-1]}(\omega)))$$

converges to

$$(x_1, x_2(\xi_1(\omega)), ..., x_T(\xi_{[T-1]}(\omega))$$

for all $\omega \in \Omega$.

Otherwise, we can just consider

$$\hat{\Omega} := \left\{\omega \in \Omega : (x_1^{(n)}, x_2^{(n)}(\xi_1(\omega)), ..., x_T^{(n)}(\xi_{[T-1]}(\omega))) \to (x_1, x_2(\xi_1(\omega)), ..., x_T(\xi_{[T-1]}(\omega))\right\},$$

which is a set of full-measure.

Now, pick $\omega \notin A$ (otherwise the inequality holds trivially), which means

$$h\left(x_1, \mathfrak{x}_2(\xi_{[1]}(\omega)), ..., \mathfrak{x}_T(\xi_{[T-1]}(\omega)), \xi_{[T]}(\omega)\right) > 0,$$

thus by the convergence almost everywhere and the lower-semicontinuity of h we have that for large enough n the following inequality is satisfied:

$$h\left(x_1^{(n)}, \mathfrak{x}_2^{(n)}(\xi_{[1]}(\omega)), ..., \mathfrak{x}_T^{(n)}(\xi_{[T]}(\omega)), \xi_{[T]}(\omega)\right) > 0,$$

which implies that $\limsup \mathbb{1}_{A_n}(\omega) = 0$, and consequently, $\limsup \mathbb{1}_{A_n}(\omega) \leq \mathbb{1}_A(\omega)$.

Finally, let us prove (4.10). Indeed, we notice that for all $n \in \mathbb{N}$

$$\begin{aligned}\varphi(x_1^{(n)}, ..., \mathfrak{x}_T^{(n)}) =& 1 - \mathbb{P}\left(-h\left(x_1^{(n)}, \mathfrak{x}_2^{(n)}(\xi_{[1]}(\omega)), ..., \mathfrak{x}_T^{(n)}(\xi_{[T]}(\omega)), \xi_{[T]}(\omega)\right) \leq 0\right) \\ &+ \mathbb{P}\left(h\left(x_1^{(n)}, \mathfrak{x}_2^{(n)}(\xi_{[1]}(\omega)), ..., \mathfrak{x}_T^{(n)}(\xi_{[T]}(\omega)), \xi_{[T]}(\omega)\right) = 0\right).\end{aligned}$$

Now, from our assumption that for all $u \in \mathbb{R}^{2T-1}$ the set $\{z \in \mathbb{R} : h(x, z) = 0\}$ has measure zero and using the fact that $\xi_{[i]}$ has density with respect to the Lebesgue measure, we have that for all $n \in \mathbb{N}$

$$\mathbb{P}\left(\{\omega \in \Omega : h\left(x_1^{(n)}, \mathfrak{x}_2^{(n)}(\xi_{[1]}(\omega)), ..., \mathfrak{x}_T^{(n)}(\xi_{[T]}(\omega)), \xi_{[T]}(\omega)\right) = 0\}\right) = 0 \qquad (4.13)$$

$$\mathbb{P}\left(\{\omega \in \Omega : h\left(x_1, \mathfrak{x}_2(\xi_{[1]}(\omega)), ..., \mathfrak{x}_T(\xi_{[T]}(\omega)), \xi_{[T]}(\omega)\right) = 0\}\right) = 0 \qquad (4.14)$$

Indeed, let us suppose that the density of $(\xi_1, \xi_2, \ldots, \xi_T)$ is f_ξ, and let us denote

$$\begin{aligned}B &:= \{\omega \in \Omega : h\left(x_1, \mathfrak{x}_2(\xi_{[1]}(\omega)), ..., \mathfrak{x}_T(\xi_{[T-1]}(\omega)), \xi_{[T]}(\omega)\right) = 0\}, \\ C &:= \{(u_1, u_2, \ldots, u_{T-1}, z) \in \mathbb{R}^T : h\left(u_1, \mathfrak{x}_2(u_1), ..., \mathfrak{x}_T(u_1, \ldots, u_{T-1}), u_1, \ldots, u_{T-1}, z\right) = 0\},\end{aligned}$$

and $u_{[T-1]} := (u_1, \ldots, u_{T-1})$. Hence,

$$\begin{aligned}\mathbb{P}(B) &= \int_{\mathbb{R}^T} f_\xi(u_{[T-1]}, z)\mathbb{1}_C(u_{[T-1]}, z)dzdu_{[T-1]} \\ &= \int_{\mathbb{R}^{T-1}} \left(\int_{\mathbb{R}} f_\xi(u_{[T-1]}, z)\mathbb{1}_C(u_{[T-1]}, z)dz\right) du_{[T-1]} = 0,\end{aligned}$$

where in the last equality we used that for all $(u_1, \ldots, u_{T-1}) \in \mathbb{R}^{T-1}$ the set

$$\{z \in \mathbb{R} : h(u_1, x_1(u_1), \ldots, \mathfrak{x}_{[T-1]}(u_{[T-1]}), u_{[T-1]}, z) = 0\}$$

has null-measure, and this in particular implies that

$$\int_{\mathbb{R}} f_{\xi}(u_{[T-1]}, z)\mathbb{1}_C(u_{[T-1]}, z)dz = 0, \text{ for all } u_{[T-1]} \in \mathbb{R}^{T-1}.$$

This yields (4.13). The proof of (4.14) follows similar arguments, so we omit the proof.

Now, let us notice that $-h$ is also lower-semicontinuous (because h is continuous), thus by the previous part

$$\begin{aligned}
\liminf \varphi(x_1^{(n)}, ..., x_T^{(n)}) = & \\
& 1 - \limsup \mathbb{P}\left(-h\left(x_1^{(n)}, x_2^{(n)}(\xi_{[1]}(\omega)), \ldots, x_T^{(n)}(\xi_{[T]}(\omega)), \xi_{[T]}(\omega)\right) \leq 0\right) \\
& \geq 1 - \mathbb{P}\left(-h\left(x_1, x_2(\xi_{[1]}(\omega)), ..., x_T(\xi_{[T-1]}(\omega)), \xi_{[T]}(\omega)\right) \leq 0\right) \\
& = \mathbb{P}\left(h\left(x_1, x_2(\xi_{[1]}(\omega)), ..., x_T(\xi_{[T-1]}(\omega)), \xi_{[T]}(\omega)\right) \leq 0\right) \\
& - \mathbb{P}\left(h\left(x_1, x_2(\xi_{[1]}(\omega)), ..., x_T(\xi_{[T-1]}(\omega)), \xi_{[T]}(\omega)\right) = 0\right) \\
& = \mathbb{P}\left(h\left(x_1, x_2(\xi_{[1]}(\omega)), ..., x_T(\xi_{[T-1]}(\omega)), \xi_{[T]}(\omega)\right) \leq 0\right),
\end{aligned}$$

which concludes the proof of (4.10).

□

The last results allows us to ensure some continuity properties of the probability function with respect to the norm and the weak-topology on $\mathcal{X}$ and $\mathcal{X}^1$, respectively. We need the following technical lemma, which in particular shows that the weak-convergence in $\mathcal{X}^1$ implies the convergence almost everywhere under subsequences. The proof of this result is based on the Rellich-Kondrachov's Theorem (see, e.g., [3, Theorem 6.3]), but with small modifications of the functions in order to avoid *regularity conditions* over the sets $\mathcal{S}_{[i]}$.

Lemma 4.14. *Consider a sequence $(x^{(n)}) \subseteq \mathcal{X}^1$ such that $x^{(n)}$ converges weakly to $x \in \mathcal{X}^1$, then there exists a subsequence $(x^{(n_k)})$, which converges almost everywhere to x.*

Proof. Consider a sequence of elements $x^{(n)} \in \mathbb{X}^1$, which converges weakly to x. Since our space $\mathcal{X}^1$ is a product spaces, it is enough to prove that each coordinates has a subsequence with the desired property. Thus, let us fix $i \in \{2, \ldots, T\}$ (the case $i = 1$ is trivial). For simpleness of the notation let us denote $f_n := x_i^{(n)}$, $f := x^{(n)}$, and $U := \mathcal{S}_{[i]}$.

Take a sequence of open bounded sets U_k such that $\bigcup_{k\in\mathbb{N}} U_k = U$ and $\overline{U_k} \subseteq U_{k+1}$ for all $k \geq 1$. Then, using a *partition of unity* (see, e.g., [119, Theorem 6.20]) we can

take functions $e_k \in C_0^\infty(U)$ such that

$$\text{a) } 0 \leq e_k(x) \leq 1, \text{ for all } x \in U, \text{ b) } e_k = 1 \text{ in } U_k, \text{ and c) } e_k = 0 \in U_{k+1}^c.$$

Since, the sets U_k are bounded we have that the inclusion $W^{r,1}(U_k) \ni u \to u \in W^{1,1}(U_k)$ is well-defined and continuous, therefore we can assume w.l.o.g. that $r = 1$. Furthermore, for each $k \geq 1$, the sequence $u_{k,n} := e_k f_n$ belongs to $W_0^{1,1}(U)$. Indeed, since the sequence $w := e_k^{1/2} f$ belongs to $W^{1,1}(U_{k+2})$ (see, e.g, [48, Theorem 1]) there exists $w_n \in C^\infty(U_{k+2})$ such that $w_n \to w$ in $W^{1,1}(U)$ (see, e.g., [3, Lemma 3.16], or [48, Theorem 2]). Hence, $e_k^{1/2} w_n \in W_0^{1,1}(U)$ (extended by zero outside of U_{k+1}) and $e_k^{1/2} w_n$ converges w in $W^{1,1}(U)$.

Now, by [3, Theorem 6.3 Part IV] we can extract a subsequence $u_{k,n_j} = e_k f_{n_j}$, which converges in norm and almost everywhere to $u_k \in L^1(U)$.

We claim that $u_k = e_k f$. Indeed, consider $v \in C_0^\infty(U)$. On the one hand by the weak convergence of f_n to f, we have

$$\lim_{j\to\infty} \int_U f_{n_j}(z) e_k(z) v(z) dz = \int_U f(z) e_k(z) v(z) dz$$

On the other hand by the norm convergence of u_{k,n_j} to u_k we have

$$\lim \int_U u_k(z) v(z) dz = \lim_{j\to\infty} \int_U u_{k,n_j}(z) v(z) dz = \lim_{j\to\infty} \int_U f_{n_j}(z) e_k(z) v(z) dz$$

Thus

$$\int_U f(z) e_k(z) v(z) dz = \lim \int_U u_k(z) v(z) dz$$

Thus, the arbitrariness of v gives us that $u_k = e_k f$ a.e. on U. In particular, for almost all $z \in U_k$ we have that

$$f_{n_j}(z) = e_k(z) f_{n_j}(z) = u_{k,n_j} \to u_k(z) = e_k(z) f(z) = f(z).$$

Therefore, using a diagonal argument we can extract a subsequence f_{n_j} which converge almost everywhere to f. □

Remark 4.15. Lemma 4.14 is not true for the Banach space $\mathcal{X}$. For example, the sequence of functions $x^n(t) = \sin(nt)$ converges weakly in $L^2[0, 2\pi]$. However, any subsquence of x^n diverges a.e. in the given space.

Now, we establish continuity properties of φ.

Theorem 4.16. *Assume that ξ has a density. Consider φ defined in* (4.9) *and assume that h is lower-semicontinous. Then φ is upper-semicontinuous on $\mathcal{X}$ with respect to the norm topology (sequentially weakly upper-semicontinuous on $\mathcal{X}^1$). Moreover, if h is continuous and for all $u \in \mathbb{R}^{2T-1}$ the set $\{z \in \mathbb{R} : h(u,z) = 0\}$ has measure zero, then φ is continuous on $\mathcal{X}$ with respect to the norm topology (sequentially weakly continuous on $\mathcal{X}^1$).*

Proof. Consider a sequence $(x_1^{(n)}, x_2^{(n)}, ..., x_T^{(n)})$ converging in $\mathcal{X}$ to $(x_1, x_2, ..., x_T) \in \mathcal{X}$, then consider a subsequence n_k such that

$$\limsup_{n\to\infty} \varphi\left(x_1^{(n_{k_j})}, x_2^{(n_{k_j})}, ..., x_T^{(n_{k_j})}\right) = \lim_{k\to\infty} \varphi\left(x_1^{(n_k)}, x_2^{(n_k)}, ..., x_T^{(n_{k-r})}\right).$$

Then, there exists a subsubsequence (n_{k_j}) such that $x_1^{(n_{k_j})} \to x_1$ and for all $i = 2, \dots, T$.

$$x_i^{(n_{k_j})}(s) \to x_i^{(n_{k_j})}(s), \text{ for almost every } s \in \mathcal{S}_{[i]}.$$

Thus, by Proposition 4.13 we have

$$\limsup_{j\to\infty} \varphi\left(x_1^{(n_{k_j})}, x_2^{(n_{k_j})}, ..., x_T^{(n_{k_j})}\right) \leq \varphi(x_1, x_2, ..., x_T)$$

Moreover,

$$\begin{aligned}\lim_{r\to\infty} \varphi\left(x_1^{(n_{k_j})}, x_2^{(n_{k_j})}, ..., x_T^{(n_{k_j})}\right) &= \lim_{k\to\infty} \varphi\left(x_1^{(n_k)}, x_2^{(n_k)}, ..., x_T^{(n_k)}\right) \\ &= \limsup_{n\to\infty} \varphi\left(x_1^{(n)}, x_2^{(n)}, ..., x_T^{(n)}\right).\end{aligned}$$

And the result follows.

Similarly, consider a subsequence $(x_1^{(n_k)}, x_2^{(n_k)}, ..., x_T^{(n_k)}) \rightharpoonup (x_1, x_2, ..., x_T) \in \mathcal{X}^1$. Then by Lemma 4.14 we can extract a subsubsequence that is pointwise a.e. convergent to $(x_1, x_2, ..., x_T)$, which, again by Proposition 4.13, ensures the sequentially weak upper semicontinuity and the sequentially weak continuity under the above hypotheses. □

4.4.3 Closedness of the Feasible Set

A second question of importance is whether the following feasible sets

$$\mathfrak{M}(p) := \{x \in \mathcal{X} : \mathbb{P}\left(h_i(x_1, ..., x_t(z_1, ..., z_{i-1}), z_1, ..., z_i) \leq 0 \ \forall i = 1, ..., T\right) \geq p\} \quad (4.15)$$

$$\mathfrak{M}^1(p) := \{x \in \mathcal{X}^1 : \mathbb{P}\left(h_i(x_1, ..., x_t(z_1, ..., z_{i-1}), z_1, ..., z_i) \leq 0 \ \forall i = 1, ..., T\right) \geq p\} \quad (4.16)$$

are closed.

Using the continuity properties of the dynamical probability function defined in (4.9) we can establish closedness of the feasible sets in $\mathcal{X}$ and in $\mathcal{X}^1$ defined above with respect to the norm topology on $\mathcal{X}$ and even with respect to the weak-topology in $\mathcal{X}^1$. Formally, we have the following results.

Formally, we have the following results.

Theorem 4.17. *Consider the probability function φ defined in* (4.9)*. Assume that h is lower semi-continuous. Then, for every $p \in [0, 1]$, the set $\mathfrak{M}(p)$ defined in* (4.15) *is strongly closed on $\mathcal{X}$, and the set $\mathfrak{M}^1(p)$ defined in* (4.16) *is sequentially weakly closed on $\mathcal{X}^1$.*

Proof. It is easy to see that $\mathfrak{M}(p) := \{x \in \mathcal{X} : \varphi(x) \geq p\}$ and $\mathfrak{M}^1(p) := \{x \in \mathcal{X}^1 : \varphi(x) \geq p\}$. Then the closedness of these sets follows directly from the upper-semicontinuity of the probability function (4.9) established in Theorem 4.16. □

4.4.4 Non-convexity of the feasible set

A third question of importance for optimization is whether the feasible set $\mathfrak{M}(p)$ is convex. In the following, we first provide a counter-example of a feasible set in $\mathcal{X}$ that is not convex in $\mathcal{X}$.

Example 4.1. *Let ξ have a normal distribution $\xi \sim \mathcal{U}([0, 4\pi] \times [0, 1])$. Let $\varphi(x) := \mathbb{P}\left(\xi_1 \leq x_1, \xi_2 \leq x_2(\xi_1)\right)$ and define a sequence $(x_1^{(n)}, x_2^{(n)})$ by*

$$x_1^{(n)} := 4\pi, \qquad x_2^{(n)} := \begin{cases} 0, & t \leq 0 \\ \sin(nt), & t \in (2\pi, 4\pi) \\ 1, & t \in [2\pi, 4\pi] \\ 0, & t > 4\pi. \end{cases}$$

Then $\mathbb{R} \times L^2(\mathbb{R}) \ni (x_1^{(n)}, x_2^{(n)}) \rightharpoonup (4\pi, \mathbb{1}_{[2\pi,4\pi]})$ and $\varphi(x_1^{(n)}, x_2^{(n)}) = \frac{1}{2} + \frac{1}{2\pi}$. However, $\varphi(4\pi, \mathbb{1}_{[2\pi,4\pi]}) = \frac{1}{2}$. Thus $x^{(n)} \in \mathfrak{M}(\frac{1}{2\pi} + 0.5)$ and $(4\pi, \mathbb{1}_{[2\pi,4\pi]}) \notin \mathfrak{M}(\frac{1}{2\pi} + 0.5)$, which shows that $\mathfrak{M}(p)$ is not necessarily sequentially weakly closed in $\mathcal{X}$.

Furthermore, suppose by contradiction that $\mathfrak{M}(p)$ is a convex set. We know by Theorem 4.17 that $\mathfrak{M}(p)$ is strongly closed, so by [29, Theorem 3.7] it is sequentially weakly closed, which, as shown above, is a contradiction.

4.5 Two-Stage Dynamic Probabilistic Constraints in $\mathcal{X}$

Analogously to the static case, we will focus on dynamic probabilistic constraints, which are linear in the random vector of the respective stage and which are two-sided and one-sided. In this section, we are interested in the properties of two-stage dynamic probabilistic constraints in the $\mathcal{X}$ space. We focus on the following three versions of the separated two-stage dynamic constraint:

$$\varphi_1(x_1, \chi_2) = \mathbb{P}\left(\begin{array}{rcl} a + x_1 \leq & \xi_1 & \leq b + x_1, \\ c + \chi_2(\xi_1) \leq & \xi_2 & \leq d + \chi_2(\xi_1) \end{array} \right), \tag{4.17}$$

or

$$\varphi_2(x_1, \chi_2) = \mathbb{P}\left(\begin{array}{rcl} a + x_1 \leq & \xi_1 & \leq b + x_1, \\ c + x_1 + \chi_2(\xi_1) \leq & \xi_2 & \leq d + x_1 + \chi_2(\xi_1) \end{array} \right), \tag{4.18}$$

or

$$\varphi_3(x_1, \chi_2) = \mathbb{P}\left(\begin{array}{rcl} a + x_1 \leq & \xi_1 & \leq b + x_1, \\ c + x_1 + \chi_2(\xi_1) - \xi_1 \leq & \xi_2 & \leq d + x_1 + \chi_2(\xi_1) - \xi_1 \end{array} \right), \tag{4.19}$$

where a, b, c, d are constant in the real extended line $[-\infty, \infty]$ with $a < b$ and $c < d$.

Firstly, we want to investigate Lipschitz continuity and differentiability of the above particular probability functions when ξ has density g_ξ. In order to examine the structural properties of the three probability functions (4.17)-(4.19), we consider the following auxiliary function $\Psi : \mathbb{R} \times \mathbb{R} \times L^2(\mathcal{S}_{[1]}) \to \mathbb{R}$ given by

$$\Psi(x, y, \chi) = \int_{a+x}^{b+x} \left(\int_{y+\chi(r)+c}^{y+\chi(r)+d} g_\xi(r, s) ds \right) dr. \tag{4.20}$$

Here, it is important to observe that

$$\varphi_1(x_1, \chi_2) = \Psi(x_1, 0, \chi_2),$$

$$\varphi_2(x_1, \chi_2) = \Psi(x_1, x_1, \chi_2).$$

Moreover, changing the random vector $\xi = (\xi_1, \xi_2)$ by $\tilde{\xi} = (\xi_1, \xi_1 + \xi_2)$ we can recover (4.19) using the results formulated for the function φ_2.

Remark 4.18. One might wish to simplify (4.20) and omit (4.18) by noticing that $\varphi_1(x_1, x_1 + x_2) = \varphi_2(x_1, x_2)$. However, the domain of φ_1 is given by $\mathbb{R} \times L^2(\mathcal{S}_1)$, so for the simple case of a Gaussian random variable, we have that the argument $x_1(\cdot) + x_2(\cdot) \notin L^2(\mathbb{R})$, since $x_1(\cdot)$ as a constant function is not integrable.

Next, we list the following properties for the probability density function; some of them will be used as assumptions to determine Lipschitz continuity and others to verify differentiability of (4.20).

Definition 4.19. Let $g_\xi : \mathbb{R}^2 \to \mathbb{R}$ be a probability density function, we say that

P(i) g_ξ is bounded if there exists $M \geq 0$ such that $g_\xi(r, s) \leq M$ for all $r, s \in \mathbb{R}$.

P(ii) g_ξ is Lipschitz continuous with respect to s uniformly in r if there exists $C \geq 0$ such that:

$$|g_\xi(r, s) - g_\xi(r, t)| \leq C|s - t|, \ \forall r, s, t \in \mathbb{R}.$$

P(iii) The marginal density of the first component of g_ξ is bounded if there exist $C \geq 0$ such that $\int_{-\infty}^{\infty} g_\xi(r, s)ds \leq C$ for all $r \in \mathbb{R}$.

P(iv) g_ξ has uniform integrable first moment provided that

$$\sup_{s \in \mathbb{R}} g_\xi(\cdot, s) \in L^1(\mathbb{R}). \tag{4.21}$$

P(v) g_ξ has uniform quadratic first moment provided that

$$\sup_{s \in \mathbb{R}} g_\xi^2(\cdot, s) \in L^2(\mathbb{R}). \tag{4.22}$$

P(vi) g_ξ is directionally Lipschitz at $\bar{r} \in \mathbb{R}$ uniformly with respect to s provided that there exist $\ell \in L^1(\mathbb{R})$ and $\epsilon > 0$ satisfying

$$\left|g_\xi(r, s) - g_\xi(\bar{r}, s)\right| \leq \ell(s)|r - \bar{r}| \ \text{ s.t. } \forall r \in [\bar{r} - \epsilon, \bar{r} + \epsilon]. \tag{4.23}$$

Let us prove that a non-degenerate Gaussian random vector satisfies our assumptions.

Lemma 4.20. *Consider $\xi \sim \mathcal{N}_2(\mu, \Sigma)$. Then its probability density satisfies Properties P(i)-P(v). Moreover, Property P(vi) is satisfied at every $\bar{r}$ in $\mathbb{R}$.*

Proof. Consider constants $C_1, C_2 > 0$ such that for every $z \in \mathbb{R}^2$,

$$-\frac{1}{2} z^\top \Sigma^{-1} z \leq -C_2 \|z\|^2, \tag{4.24}$$

$$\|\Sigma^{-1} z\| \leq \|z\| C_1. \tag{4.25}$$

Pick $C = \frac{1}{2\pi\sqrt{|\Sigma|}}$. We have that

$$\begin{aligned} g_\xi(r,s) &= C \exp\left(-\frac{1}{2} (r-\mu_1, s-\mu_2) \Sigma^{-1} \begin{pmatrix} r-\mu_1 \\ s-\mu_2 \end{pmatrix} \right) \\ (\text{recall (4.24)}) &\leq C \exp(-C_2(r-\mu_1)^2 - C_2(s-\mu_2)^2) \\ &\leq C h_1(r) h_2(s) \qquad (4.26) \\ &\leq C. \qquad (4.27) \end{aligned}$$

where $h_1(r) := \exp(-C_2(r-\mu_1)^2) \leq 1$ and $h_2(s) := \exp(-C_2(s-\mu_2)^2) \leq 1$. Thus, P(i) is satisfied. In addition, it also follows by (4.26) that

$$\int_{-\infty}^{\infty} g_\xi(r,s) ds \leq \int_{-\infty}^{\infty} C h_2(s) ds = \frac{C\sqrt{\pi}}{C_2} \quad \text{for all} \quad r \in \mathbb{R},$$

so P(iii) is also satisfied. Moreover, since g_ξ is bounded and by (4.26) we have that for every $\varkappa_2 \in L^2(\mathbb{R})$,

$$g_\xi(r, \varkappa_2(r)) \leq \sup_{s \in \mathbb{R}} g(r,s) \leq C h_1(r).$$

Due to $\int_{-\infty}^{\infty} C \cdot h_1(r) dr < \infty$ and $\int_{-\infty}^{\infty} C^2 \cdot h_1^2(r) dr < \infty$, P(iv) and P(v) hold true. Now, let us compute the gradient of g_ξ, which is given by

$$\nabla g_\xi(r,z) = -g_\xi(r,s)\big(r-\mu_1, s-\mu_2\big)\Sigma^{-1}$$

(since Σ^{-1} is a symmetric matrix). We claim that there exists some constant $\tilde{C}$ such that

$$\|\nabla g_\xi(r,s)\| \leq \tilde{C} h_1^{1/2}(r) h_2^{1/2}(s) \quad \forall r, s \in \mathbb{R}. \tag{4.28}$$

Indeed, by (4.25) and (4.26), it holds that

$$\|\nabla g_{\tilde{\xi}}(r,s)\| \leq \|g_{\tilde{\xi}}(r,s)\| \cdot \|(r,s) - (\mu_1,\mu_2)\| \cdot \|\Sigma^{-1}\| \tag{4.29}$$

$$\leq C \cdot C_1 h_1(r) h_2(s) \|(r,s) - (\mu_1,\mu_2)\| \tag{4.30}$$

$$= C \cdot C_1 \sqrt{h_1(r)h_2(s)h_1(r)h_2(s)} \|(r,s) - (\mu_1,\mu_2)\|. \tag{4.31}$$

Now let us notice that the function

$$\gamma(r,s) := h_1^{1/2}(r) h_2^{1/2}(s) \|(r,s) - (\mu_1,\mu_2)\|$$

attains its maximum, consequently, let us denote $\tilde{C} := C \cdot C_1 \cdot (\max_{(r,s)\in\mathbb{R}^2} \gamma(r,s))$, using this constant we have that (4.28) holds. Hence, $\nabla g_{\tilde{\xi}}$ is bounded. So, by the Mean Value Theorem, we obtain that for any $r \in \mathbb{R}$ (recall (4.28)) there exists some $\tilde{r} \in [\bar{r}, r]$ such that

$$|g_{\tilde{\xi}}(r,s) - g_{\tilde{\xi}}(\bar{r},s)| \leq \|\nabla g_{\tilde{\xi}}(\tilde{r},s)\| \cdot |r - \bar{r}| \leq \tilde{C} h_2^{1/2}(s) |r - \bar{r}|, \tag{4.32}$$

which implies that P(vi) holds. Moreover, since $h_2^{1/2}(s) \leq 1$, it follows by (4.32) that P(ii) also holds. □

Observation 4.5.1. *It is worth mentioning that if a random vector $\xi = (\xi_1, \xi_2)$ satisfies one of the properties P(i)-P(v), then $\tilde{\xi} = (\xi_1, \xi_1 + \xi_2)$ satisfies the same property. Moreover, it is well-known that if $\xi \sim \mathcal{N}((\mu_1,\mu_2), \Sigma)$, then*

$$\tilde{\xi} \sim \mathcal{N}\left((\mu_1, \mu_1 + \mu_2), \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \Sigma \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\right).$$

Let $\xi = (\xi_1, \xi_2)$ be a random vector with density g_ξ. Then, the random vector $\tilde{\xi} := (\xi_1, \xi_1 + \xi_2)$ has density $g_{\tilde{\xi}}(r,s) := g_\xi(r, r+s)$. Moreover, if ξ satisfies one of the properties P(i)-P(v), then $\tilde{\xi}$ also satisfies the same property.

Additionally, we prove that a uniform random vector satisfies some of our assumptions.

Lemma 4.21. *Let $A := [a,b] \times [c,d] \subset \mathbb{R}^2$ and consider $\xi \sim \mathcal{U}(A)$. Then its probability density satisfies Properties P(i), P(iii), P(iv) and P(v).*

Proof. Let $c := area(A)$, then we have that

$$g_\xi(r,s) = \frac{1}{c} \mathbf{1}_{\{(r,s)\in A\}} \leq 1 \text{ for all } r,s \in \mathbb{R},$$

so it satisfies P(i). It directly follows that $\int_{-\infty}^{\infty} g_{\xi}(r,s)ds = \int_A \frac{1}{c} ds = 1$ for all $r \in \mathbb{R}$, so P(iii) holds. Lastly, since g_{ξ} is bounded, we can deduce that for every $\varkappa_2 \in L^2(\mathbb{R})$

$$\int_{-\infty}^{\infty} g_{\xi}(r, \varkappa_2(r))dr \leq \int_{-\infty}^{\infty} \sup_{s\in\mathbb{R}} g(r,s)dr \leq \int_A dr = c < +\infty,$$

such that P(iv) is satisfied. Likewise, $\int_{-\infty}^{\infty} \sup_{s\in\mathbb{R}} g_{\xi}^2(r,s)dr \leq c^2 < +\infty$, so P(v) also holds true. □

4.5.1 Lipschitz-Continuity of Ψ

We start this subsection, showing that a broad class of density distribution, which includes the non-degenerate Gaussian distributions and the uniform distribution, satisfies that the probability function (4.20) is Lipschitz continuous.

Lemma 4.22. *Let $\xi := (\xi_1, \xi_2)$ be a random vector with density g_{ξ}. Consider $x_1, x_2, y_1, y_2 \in \mathbb{R}$ and $\varkappa_1, \varkappa_2 \in L^2(S_i)$. Then the following holds true:*

1. *If ξ satisfies P(iii), we have that*

$$|\Psi(x_1, y_1, \varkappa_1) - \Psi(x_2, y_1, \varkappa_1)| \leq M_1 |x_1 - x_2|, \tag{4.33}$$

where $M_1 := 2 \sup_{r\in\mathbb{R}} \int_{-\infty}^{\infty} g(s,r)ds.$

2. *If ξ satisfies P(iv), we have that*

$$|\Psi(x_1, y_1, \varkappa_1) - \Psi(x_1, y_2, \varkappa_1)| \leq M_2 |y_1 - y_2|, \tag{4.34}$$

where $M_2 := 2 \int_{-\infty}^{\infty} \sup_{s\in\mathbb{R}} g(s,r)dr.$

3. *If ξ satisfies P(v), we have*

$$|\Psi(x_1, y_1, \varkappa_1) - \Psi(x_1, y_1, \varkappa_2)| \leq M_3 \|\varkappa_1 - \varkappa_2\|,$$

where $M_3 := 2 \left(\int_{-\infty}^{\infty} \sup_{s\in\mathbb{R}} g^2(s,r)dr \right)^{1/2}.$

Proof. Suppose w.l.o.g. that $x_1 \leq x_2$ and $y_1 \leq y_2$. Then

$$
\begin{aligned}
|\Psi(x_1, y_1, \mathfrak{x}_1) - \Psi(x_2, y_1, \mathfrak{x}_1)| &= \left| \int_{a+x_1}^{b+x_1} \int_{c+y_1+\mathfrak{x}_1(r)}^{d+y_1+\mathfrak{x}_1(r)} g_{\tilde{\xi}}(r,s) ds dr \right. \\
&\qquad \left. - \int_{a+x_2}^{b+x_2} \int_{c+y_1+\mathfrak{x}_1(r)}^{d+y_1+\mathfrak{x}_1(r)} g_{\tilde{\xi}}(r,s) ds dr \right| \\
&\leq \left| \int_{a+x_1}^{b+x_1} \int_{c+y_1+\mathfrak{x}_1(r)}^{d+y_1+\mathfrak{x}_1(r)} g_{\tilde{\xi}}(r,s) ds dr \right. \\
&\qquad \left. - \int_{a+x_2}^{b+x_1} \int_{c+y_1+\mathfrak{x}_1(r)}^{d+y_1+\mathfrak{x}_1(r)} g_{\tilde{\xi}}(r,s) ds dr \right| \\
&\qquad + \left| \int_{a+x_2}^{b+x_1} \int_{c+y_1+\mathfrak{x}_1(r)}^{d+y_1+\mathfrak{x}_1(r)} g_{\tilde{\xi}}(r,s) ds dr \right. \\
&\qquad \left. - \int_{a+x_2}^{b+x_2} \int_{c+y_1+\mathfrak{x}_1(r)}^{d+y_1+\mathfrak{x}_1(r)} g_{\tilde{\xi}}(r,s) ds dr \right| \\
&\leq \left| \int_{a+x_1}^{a+x_2} \int_{-\infty}^{+\infty} g_{\tilde{\xi}}(r,s) ds dr \right| \\
&\qquad + \left| \int_{b+x_1}^{b+x_2} \int_{-\infty}^{+\infty} g_{\tilde{\xi}}(r,s) ds dr \right| \\
&\leq 2 \left(\sup_{r \in \mathbb{R}} \int_{-\infty}^{+\infty} g_{\tilde{\xi}}(r,s) ds \right) |x_1 - x_2| \\
&= M_1 |x_1 - x_2|.
\end{aligned}
$$

$$
\begin{aligned}
|\Psi(x_1,y_1,x_{\mathfrak{1}})-\Psi(x_1,y_2,x_{\mathfrak{1}})| =& \left| \int_{a+x_1}^{b+x_1} \int_{c+y_1+x_{\mathfrak{1}}(r)}^{d+y_1+x_{\mathfrak{1}}(r)} g_{\xi}(r,s)dsdr \right. \\
& \left. - \int_{a+x_1}^{b+x_1} \int_{c+y_2+x_{\mathfrak{1}}(r)}^{d+y_2+x_{\mathfrak{1}}(r)} g_{\xi}(r,s)dsdr \right| \\
\leq & \left| \int_{a+x_1}^{b+x_1} \left(\int_{c+y_1+x_{\mathfrak{1}}(r)}^{d+y_1+x_{\mathfrak{1}}(r)} g_{\xi}(r,s)ds - \int_{c+y_2+x_{\mathfrak{1}}(r)}^{d+y_1+x_{\mathfrak{1}}(r)} g_{\xi}(r,s)ds \right) dr \right| \\
& + \left| \int_{a+x_1}^{b+x_1} \left(\int_{c+y_2+x_{\mathfrak{1}}(r)}^{d+y_1+x_{\mathfrak{1}}(r)} g_{\xi}(r,s)ds \right. \right. \\
& \left. \left. - \int_{c+y_2+x_{\mathfrak{1}}(r)}^{d+y_2+x_{\mathfrak{1}}(r)} g_{\xi}(r,s)ds \right) dr \right| \\
\leq & \left| \int_{a+x_1}^{b+x_1} \int_{c+y_1+x_{\mathfrak{1}}(r)}^{c+y_2+x_{\mathfrak{1}}(r)} g_{\xi}(r,s)ds \right| \\
& + \left| \int_{a+x_1}^{b+x_1} \int_{d+y_1+x_{\mathfrak{1}}(r)}^{d+y_2+x_{\mathfrak{1}}(r)} g_{\xi}(r,s)ds \right| \\
\leq & 2 \left(\int_{-\infty}^{+\infty} \sup_{s\in\mathbb{R}} g_{\xi}(r,s)dr \right) |y_1-y_2| \\
= & M_2|y_1-y_2|.
\end{aligned}
$$

$$
\begin{aligned}
|\Psi(x_1, y_1, \varkappa_1) - \Psi(x_1, y_1, \varkappa_2)| &= \left| \int_{a+x_1}^{b+x_1} \int_{c+y_1+\varkappa_1(r)}^{d+y_1+\varkappa_1(r)} g_\xi(r,s)dsdr \right. \\
&\qquad \left. - \int_{a+x_1}^{b+x_1} \int_{c+y_1+\varkappa_2(r)}^{d+y_1+\varkappa_2(r)} g_\xi(r,s)dsdr \right| \\
&\leq \left| \int_{a+x_1}^{b+x_1} \int_{c+y_1+\varkappa_1(r)}^{d+y_1+\varkappa_1(r)} g_\xi(r,s)dsdr \right. \\
&\qquad \left. - \int_{a+x_1}^{b+x_1} \int_{c+y_1+\varkappa_2(r)}^{d+y_1+\varkappa_1(r)} g_\xi(r,s)dsdr \right| \\
&\qquad + \left| \int_{a+x_1}^{b+x_1} \int_{d+y_1+\varkappa_1(r)}^{c+y_1+\varkappa_2(r)} g_\xi(r,s)dsdr \right. \\
&\qquad \left. - \int_{a+x_1}^{b+x_1} \int_{d+y_1+\varkappa_1(r)}^{c+y_1+\varkappa_2(r)} g_\xi(r,s)dsdr \right| \\
&\leq \int_{a+x_1}^{b+x_1} \left| \int_{c+y_1+\varkappa_1(r)}^{c+y_1+\varkappa_2(r)} g_\xi(r,s)ds \right| dr \\
&\qquad + \int_{a+x_1}^{b+x_1} \left| \int_{d+y_1+\varkappa_2(r)}^{d+y_1+\varkappa_1(r)} g_\xi(r,s)ds \right| dr \\
&\leq \int_{a+x_1}^{b+x_1} \sup_{s\in\mathbb{R}} g_\xi(r,s)|\varkappa_1(r) - \varkappa_2(r)|dr \\
&\qquad + \int_{a+x_1}^{b+x_1} \sup_{s\in\mathbb{R}} g_\xi(r,s)|\varkappa_1(r) - \varkappa_2(r)|dr \\
&\leq \int_{-\infty}^{+\infty} \sup_{s\in\mathbb{R}} g_\xi(r,s)|\varkappa_1(r) - \varkappa_2(r)|dr \\
&\qquad + \int_{-\infty}^{+\infty} \sup_{s\in\mathbb{R}} g_\xi(r,s)|\varkappa_1(r) - \varkappa_2(r)|dr \\
\text{(by Cauchy-Schwartz inequality)} \quad &\leq 2\left(\int_{-\infty}^{+\infty} \sup_{s\in\mathbb{R}} g_\xi^2(r,s) \right)^{1/2} \|\varkappa_1 - \varkappa_2\| \\
&= M_3\|\varkappa_1 - \varkappa_2\|.
\end{aligned}
$$

□

Theorem 4.23. *[Lipschitz-Continuity of Ψ] Suppose ξ has probability density g_ξ. If g_ξ satisfies properties P(iii) and P(v), we have that the probability function φ_1 defined in* (4.17) *is Lipschitz continuous. In addition, if g_ξ satisfies property $P(iv)$, the probability function φ_2 defined in* (4.18) *is Lipschitz continuous.*

Proof. Follows directly from Lemma 4.22 and the corresponding calculations under the given assumptions. □

In particular, the above result allows us to write the following statement for a non-degenerate Gaussian vector.

Corollary 4.24. *Assume that ξ is a non-degenerate Gaussian vector. Then the probability functions defined in* (4.17), (4.18) *and* (4.19) *are Lipschitz continuous.*

Likewise, Theorem 4.23 and Lemma 4.21 allow us to write the following statement for a uniform vector.

Corollary 4.25. *Assume that ξ is a uniform random vector with compact support $[a, b] \times [c, d] \subset \mathbb{R}^2$. Then the probability functions defined in* (4.17), (4.18) *are Lipschitz continuous.*

4.5.2 Differentiability of Ψ

In our models (4.17), (4.18) and (4.19) we ask: when are they differentiable? The following example shows that this property cannot be ensured at every pair of points $(x_1, \varkappa_2)$.

Example 4.2. *Suppose that $\xi = (\xi_1, \xi_2) \sim \mathcal{N}_2(0, \mathbb{I})$. Moreover, let the probability function $\varphi : \mathbb{R} \times L^2(\mathbb{R}) \to \mathbb{R}$ be given by*

$$\varphi(x_1, \varkappa_1) = \mathbb{P}(\xi_1 \leq x_1, \xi_2 \leq \varkappa_1(\xi_1)),$$

and let us fix $\varkappa_2 := \mathbb{1}_{[0,1]}$.
Then, $\varphi(\cdot, \varkappa)$ (and thus φ) fails to be differentiable. Indeed

$$\varphi(\varkappa_1, \varkappa_2) = \begin{cases} \Phi(x_1)\Phi(0) & \text{if } x_1 < 0 \\ \Phi(0)\Phi(0) + (\Phi(x_1) - \Phi(0))\Phi(1) & \text{if } x_1 \in [0, 1] \\ \Phi(0)\Phi(0) + (\Phi(1) - \Phi(0))\Phi(1) + (\Phi(x_1) - \Phi(1))\Phi(0) & \text{if } x_1 > 1, \end{cases}$$

where Φ is the standard Gaussian cumulative distribution.

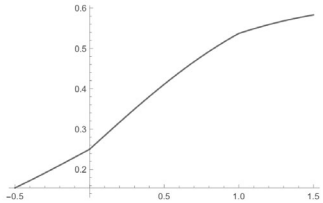


FIGURE 4.2: Illustration of non-differentiable $\varphi(\cdot, \varkappa_2)$ with $\varkappa_2$ discontinuos.

The above example motivates us to study differentiablity at point (x, χ) at which χ is continuous. Actually, we will see that under this assumption (among others), the partial derivative exists. Before presenting our main results, which are presented in Theorem 4.30 and Corollary 4.31, we need to establish auxiliary lemmas about the partial differentiability with respect to each variable.

The next lemma is just a technical result.

Lemma 4.26. *Let $g_\xi : \mathbb{R}^2 \to \mathbb{R}$ be a probability density function that satisfies property P(vi) at every $\bar{r} \in \mathbb{R}$. Assume that $\chi : \mathbb{R} \to \mathbb{R}$ is a continuous function. Then the function*

$$\alpha(r) := \int_{c+y+\chi(r)}^{d+y+\chi(r)} g_\xi(r, s)ds \tag{4.35}$$

is continuous.

Proof. Let $\{r_n\}_{n\in\mathbb{N}}$ be a sequence of real numbers such that r_n converges to $\bar{r}$ with $|r_n - \bar{r}| \leq 1$. Then it holds true that

$$\begin{aligned}\lim_{n\to\infty} \alpha(r_n) &= \lim_{n\to\infty} \int_{c+y+\chi(r_n)}^{d+y+\chi(r_n)} g_\xi(r_n, s)ds \\ &= \lim_{n\to\infty} \int_{-\infty}^{\infty} g_\xi(r_n, s)\mathbf{1}_{[c+y+\chi(r_n),d+y+\chi(r_n)]}(s)ds.\end{aligned}$$

Now, let us notice that

$$\lim_{n\to\infty} g_\xi(r_n, s)\mathbf{1}_{[c+y+\chi(r_n),d+y+\chi(r_n)]}(s) = g_\xi(\bar{r}, s)\mathbf{1}_{[c+y+\chi(\bar{r}),d+y+\chi(\bar{r})]}(s) \;\; a.e. \tag{4.36}$$

Indeed, there are three cases. If $s \in (c + y + \chi(\bar{r}), d + y + \chi(\bar{r}))$ then by continuity of χ_2 in $\bar{r}$ we have that $s \in (c + y + \chi(r_n), d + y + \chi(r_n))$ for large enough n, so by continuity of g at $\bar{r}$,

$$\lim_{n\to\infty} g_\xi(r_n, s)\mathbf{1}_{[c+y+\chi(r_n),d+y+\chi(r_n)]}(s) = g_\xi(\bar{r}, s)\mathbf{1}_{[c+y+\chi(r),d+y+\chi(r)]}(s).$$

If $s \notin [c + y + \chi(\bar{r}), d + y + \chi(\bar{r})]$, the result follows analogously. Lastly, if $s = c + y + \chi(\bar{r})$, or $s = d + y + \chi(\bar{r})$ we cannot assure that $s \in [c + y + \chi(\bar{r}_n), d + y + \chi(\bar{r}_n)]$for large enough n. However $\lambda(\{c + y + \chi(\bar{r}), d + y + \chi(\bar{r})\}) = 0$. Therefore, (4.36) holds.

Moreover, by (4.23) it follows that (recall $|r_n - \bar{r}| \leq 1$)

$$\begin{aligned}\left|g_\xi(r_n, s)\mathbf{1}_{[c+y+\chi(r_n),d+y+\chi(r_n)]}(s) - g_\xi(\bar{r}, s)\mathbf{1}_{[c+y+\chi(\bar{r}),d+y+\chi(\bar{r})]}(s)\right| &\leq |g(r_n, s) - g(\bar{r}, s)| \\ &\leq \ell(s)|r_n - \bar{r}| \leq \ell(s).\end{aligned}$$

So, we further conclude

$$g_\xi(r_n, s)\mathbf{1}_{[c+y+\chi(r_n),d+y+\chi(r_n)]}(s) \leq \ell(s) + g(\bar{r}, s).$$

We finally note that $\ell(\cdot) + g(\bar{r}, \cdot) \in L^1(\mathbb{R})$ and using (4.36) we can apply Lebesgue's Dominated Convergence Theorem so that $g_\xi(\bar{r}, s)\mathbf{1}_{(-\infty,\chi(\bar{r})]}(s)$ is integrable and

$$\begin{aligned}\lim_{n\to\infty}\int_{-\infty}^{\infty} g_\xi(r_n, s)\mathbf{1}_{[c+y+\chi(r_n),d+y+\chi(r_n)]}(s)ds &= \int_{-\infty}^{\infty}\lim_{n\to\infty} g_\xi(r_n, s)\mathbf{1}_{[c+y+\chi(r_n),d+y+\chi(r_n)]}(s)ds \\ &= \int_{-\infty}^{\infty} g_\xi(\bar{r}, s)\mathbf{1}_{[c+y+\chi(\bar{r}),d+y+\chi(\bar{r})]}(s)ds \\ &= \alpha(\bar{r})\end{aligned}$$

holds. □

In the next lemma, we present the partial differentiability of Ψ with respect to the first decision variable.

Lemma 4.27. *Under the assumptions of Lemma 4.26, we have that Ψ is partially differentiable with respect to x and*

$$\frac{\partial\Psi(x, y, \chi)}{\partial x} = \int_{c+y+\chi(x)}^{d+y+\chi(x)} g_\xi(x, s)ds. \tag{4.37}$$

Proof. First we notice that

$$\Psi(x, y, \chi) = \int_{a+x}^{b+x} \alpha(r)dr, \tag{4.38}$$

where α is defined in (4.35). Now, by Lemma 4.26 we know that α is continuous, thus by the *fundamental theorem of calculus* (see, e.g., [35, Theorem 6.3.3] or [6]) we get that (4.37) holds. □

We continue with the gradient of our probability function with respect to χ.

Lemma 4.28. *Assume that ξ satisfies properties P(i), P(ii) and P(v). Then the function defined in* (4.20) *is partially differentiable with respect to χ at every $(x, \chi) \in \mathbb{R} \times L^2(S_{[1]})$ and its partial gradient is given by*

$$\nabla_\chi\Psi(x, y, \chi) = g_\xi\left(\cdot, d + y + \chi(\cdot)\right)\mathbf{1}_{[a+x,b+x]}(\cdot) - g_\xi\left(\cdot, c + y + \chi(\cdot)\right)\mathbf{1}_{[a+x,b+x]}(\cdot), \tag{4.39}$$

with the convention $g_\xi(\cdot, -\infty) = 0 = g_\xi(\cdot, +\infty)$, if one of the extreme point c or d is equal to $-\infty$ or $+\infty$, respectively.

Proof. By applying a convenient substitution, w.l.o.g., we may assume that $x = y = 0$. Moreover, from the fact that

$$\int_a^b \int_{c+\chi(r)}^{d+\chi(r)} g_\xi(r,s)drds = \int_a^b \int_{-\infty}^{d+\chi(r)} g_\xi(r,s)drds - \int_a^b \int_{-\infty}^{c+\chi(r)} g_\xi(r,s)drds$$

and the symmetry of the involved functions, we only need to compute the partial derivative of

$$\tilde{\varphi}(\chi) = \int_a^b \int_{-\infty}^{d+\chi(r)} g_\xi(r,s)drds.$$

Let h be an arbitrary function in $L^2(\mathcal{S}_{[1]})$ and let

$$A(h) := \int_a^b g(r, d+\chi(r))h(r)dr.$$

We note that $A(h)$ is a linear function on h, and by property P(v) it follows that

$$\left(\int_{-\infty}^{+\infty} g_\xi^2(r, d+\chi(r))\right)^{1/2} < \infty.$$

Thus,

$$\begin{aligned} |A(h)| &\leq \int_{-\infty}^{\infty} |g_\xi(r,\chi(r))||h(r)|dr \\ &\leq \left(\int_{-\infty}^{\infty} g_\xi(r, d+\chi(r))^2 dr\right)^{1/2} \left(\int_{-\infty}^{\infty} h(r)^2 dr\right)^{1/2} \\ &\quad \text{(by Cauchy-Schwarz's inequality).} \end{aligned}$$

The last inequality provides the continuity of A.

Let us denote by

$$\kappa := \lim_{||h||_2 \to 0} \frac{1}{||h||_2} \left|\tilde{\varphi}(\chi) - \tilde{\varphi}(\chi) - A(h)\right|.$$

We will show that this limit exists and is equal to zero. Indeed,

$$\begin{aligned}\kappa &= \lim_{||h||_2\to 0}\frac{1}{||h||_2}\left|\tilde{\varphi}(\chi+h)-\tilde{\varphi}(\chi)-\int_a^b g(r,d+\chi(r))\cdot h(r)dr\right| \\ &\quad \lim_{||h||_2\to 0}\frac{1}{||h||_2}\left|\int_a^b\int_{-\infty}^{d+\chi(r)+h(r)} g_{\xi}(r,s)dsdr-\int_a^b\int_{-\infty}^{d+\chi(r)} g_{\xi}(r,s)dsdr\right. \\ &\quad \left.-\int_a^b g_{\xi}(r,d+\chi(r))\cdot h(r)dr\right| \\ &= \lim_{||h||_2\to 0}\frac{1}{||h||_2}\left|\int_a^b\int_{d+\chi(r)}^{d+\chi(r)+h(r)} g_{\xi}(r,s)dsdr-\int_a^b g_{\xi}(r,d+\chi(r))\cdot h(r)dr\right|.\end{aligned}$$

Moreover, for every $r\in\mathbb{R}$ we have

$$g_{\xi}(r,d+\chi(r))h(r)=\int_{d+\chi(r)}^{d+x_2(r)+h(r)} g(r,d+\chi(r))ds,$$

consequently

$$\kappa=\lim_{||h||_2\to 0}\frac{1}{||h||_2}\left|\int_a^b\int_{\chi(r)}^{\chi(r)+h(r)}(g_{\xi}(r,s)-g(r,d+\chi(r))dsdr\right|.$$

Moreover, by property P(ii) we have that (for some constant $C>0$)

$$|g_{\xi}(r,s)-g_{\xi}(r,d+\chi(r))|\leq C|s-(d+\chi(r))|\leq C|h(r)|$$

for all $r\in\mathbb{R}$ and $s\in[d+\chi(r),d+\chi(r)+h(r)]$.

Hence,

$$\begin{aligned}\kappa &\leq \lim_{||h||_2\to 0}\frac{1}{||h||_2}\left|\int_a^b C|h(r)|\cdot|h(r)|dr\right| \\ &\leq \lim_{||h||_2\to 0}\frac{1}{||h||_2}\cdot C\cdot||h||_2^2=0.\end{aligned}$$

Therefore, we conclude that $\tilde{\varphi}$ is Fréchet differentiable with respect to χ and consequently we have the desired result. □

Proposition 4.29. *Under the assumption of Lemma 4.28, the partial derivative of the function* (4.20) *with respect to* χ *is continuous on* $\mathbb{R}\times\mathbb{R}\times L^2(S_{[1]})$.

Proof. By Lemma 4.28 we have that

$$\nabla_{\chi}\Psi(x,y,\chi)=g_{\xi}\left(\cdot,d+y+\chi(\cdot)\right)\mathbf{1}_{[a+x,b+x]}(\cdot)-g_{\xi}\left(\cdot,c+y+\chi(\cdot)\right)\mathbf{1}_{[a+x,b+x]}(\cdot).$$

Now, consider a sequence

$$\mathbb{R} \times \mathbb{R} \times L^2(\mathbb{R}) \ni (x_n, y_n \varkappa_n) \to \varkappa = (x, y, \varkappa).$$

Then by passing to a subsequence, we may assume that $\varkappa_n \to \varkappa$ almost everywhere, hence by the continuity of g_ξ with respect to the second variable (recall property P(ii)), we have that for almost all $u \in \mathbb{R}$,

$$\nabla_\varkappa \Psi(x_n, y_n, \varkappa_n)(u) \to \nabla_\varkappa \Psi(x, y, \varkappa)(u).$$

Now, we notice that for all $r \in \mathbb{R}$,

$$|\nabla_\varkappa \Psi(x_n, y_n, \varkappa_n)(r)| \leq 2 \sup_{s \in \mathbb{R}} g_\xi(r, s) := \nu(r),$$

and by property P(v), ν belongs to $L^2(\mathbb{R})$.

Therefore, by Lebesgue's dominated convergence theorem we have our desired conclusion.

□

Now, using the above lemmas, let us write a general statement for the partial differentiability of probability functions (4.17), (4.18) and (4.19).

Theorem 4.30. *Consider a random vector ξ satisfying properties P(i) to P(vi). Then the probability functions* (4.17) *and* (4.18) *are continuously partially differentiable with respect to $\varkappa$ at any $(x, \varkappa) \in \mathbb{R} \times L^2(S_{[i]})$, and the following formulae hold:*

$$\nabla_\varkappa \varphi_1(x, \varkappa) = g_\xi(\cdot, d + \varkappa(\cdot)) \mathbf{1}_{[a+x, b+x]} - g_\xi(\cdot, c + \varkappa(\cdot)) \mathbf{1}_{[a+x, b+x]}(\cdot),$$
$$\nabla_\varkappa \varphi_2(x, \varkappa) = g_\xi(\cdot, d + x + \varkappa(\cdot)) \mathbf{1}_{[a+x, b+x]} - g_\xi(\cdot, c + x + \varkappa(\cdot)) \mathbf{1}_{[a+x, b+x]}(\cdot),$$

respectively, with the convention that $g_\xi(r, s) = 0$ if s, r do not belong to $\mathbb{R}$. In addition, if $\varkappa$ is continuous, then the probability functions (4.17) *and* (4.18) *are also partially differentiable with respect to x, and their partial derivatives are given by*

$$\frac{\partial \varphi_1(x, \varkappa)}{\partial x} = \int_{c+\varkappa(x)}^{d+\varkappa(x)} g_\xi(x, s),$$
$$\frac{\partial \varphi_2(x, \varkappa)}{\partial x} = \int_{c+x+\varkappa(x)}^{d+x+\varkappa(x)} g_\xi(x, s),$$

respectively.

Proof. The first part corresponds to an application of Lemma 4.28 and Proposition 4.29, and the second part corresponds to an application of Lemma 4.27. □

Next, we write explicit formulae for the case of a non-degenerate Gaussian random vector.

Corollary 4.31. *Assume that $\xi \sim (\mu, \Sigma)$ is a non-degenerate Gaussian random vector. Then the probability functions* (4.17) *and* (4.18) *are (continuously) partially differentiable with respect to χ at any $(x, \chi) \in \mathbb{R} \times L^2(\mathbb{R})$, and the following formulae hold:*

$$\begin{aligned}\nabla_\chi \varphi_1(x,\chi)(u) &= \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\Bigg(-\frac{1}{2}\Big(u-\mu_1, d+\chi(u)-\mu_2\Big)\Sigma^{-1} \\ &\quad \begin{pmatrix} u-\mu_1 \\ d+\chi(u)-\mu_2 \end{pmatrix}\Bigg)\mathbf{1}_{[a+x,b+x]}(u) \\ &\quad -\frac{1}{2\pi\sqrt{|\Sigma|}} \exp\Bigg(-\frac{1}{2}\Big(u-\mu_1, c+\chi(u)-\mu_2\Big) \\ &\quad \Sigma^{-1}\begin{pmatrix} u-\mu_1 \\ c+\chi(u)-\mu_2 \end{pmatrix}\Bigg)\mathbf{1}_{[a+x,b+x]}(u).\end{aligned}$$

$$\begin{aligned}\nabla_\chi \varphi_2(x,\chi)(u) &= \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\Bigg(-\frac{1}{2}\Big(u-\mu_1, d+x+\chi(u)-\mu_2\Big)\Sigma^{-1} \\ &\quad \begin{pmatrix} u-\mu_1 \\ d+x+\chi(u)-\mu_2 \end{pmatrix}\Bigg)\mathbf{1}_{[a+x,b+x]}(u) \\ &\quad -\frac{1}{2\pi\sqrt{|\Sigma|}} \exp\Bigg(-\frac{1}{2}\Big(u-\mu_1, c+x+\chi(u)-\mu_2\Big) \\ &\quad \Sigma^{-1}\begin{pmatrix} u-\mu_1 \\ c+x+\chi(u)-\mu_2 \end{pmatrix}\Bigg)\mathbf{1}_{[a+x,b+x]}(u).\end{aligned}$$

In addition, if χ is continuous, then the probability functions (4.17) *and* (4.18) *are also partially differentiable with respect to x, and their partial derivatives are given by*

$$\begin{aligned}\frac{\partial \varphi_1(x,\chi)}{\partial x} &= \frac{1}{2\pi\sqrt{|\Sigma|}} \int_{c+\chi(x)}^{d+\chi(x)} \exp\Bigg(-\frac{1}{2}\Big(x-\mu_1, s-\mu_2\Big)\Sigma^{-1}\begin{pmatrix} x-\mu_1 \\ s-\mu_2 \end{pmatrix}\Bigg) ds \\ \frac{\partial \varphi_2(x,\chi)}{\partial x} &= \frac{1}{2\pi\sqrt{|\Sigma|}} \int_{c+x+\chi(x)}^{d+x+\chi(x)} \exp\Bigg(-\frac{1}{2}\Big(x-\mu_1, s-\mu_2\Big)\Sigma^{-1}\begin{pmatrix} x-\mu_1 \\ s-\mu_2 \end{pmatrix}\Bigg) ds\end{aligned}$$

Proof. The formulae for φ_1 and φ_2 follow directly from Theorem 4.30. □

An important special case of the linear separated model (4.17) arises when we have the one-sided probability function

$$\hat{\varphi}_1(x, \chi) := \mathbb{P}(\xi_1 \leq x, \xi_2 \leq \chi(\xi_1)), \tag{4.40}$$

with $a = c = -\infty$ and $b = d = 0$. Lastly, for the Gaussian case, we present an explicit formula for the partial derivative of (4.40).

Corollary 4.32. *Assume that $\xi \sim \mathcal{N}(\mu, \Sigma)$ is a non-degenerate Gaussian random vector with*

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

Then the one-sided probability function φ_1 from (4.17) *with $a = c = -\infty$ and $b = d = 0$, denoted by $\hat{\varphi}_1$, is partially differentiable with respect to χ at any $(x, \chi) \in \mathbb{R} \times L^2(\mathbb{R})$, and the following formula holds:*

$$\nabla_\chi \hat{\varphi}_1(x, \chi)(r) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}\Big(r - \mu_1, \chi(r) - \mu_2\Big)\Sigma^{-1}\begin{pmatrix} r - \mu_1 \\ \chi(r) - \mu_2 \end{pmatrix}\right)\mathbf{1}_{(-\infty, x]}(r). \tag{4.41}$$

In addition, if χ is continuous, then the probability function (4.20) *is also partially differentiable with respect to x, and its partial derivative is given by*

$$\frac{\partial \hat{\varphi}_1}{\partial x}(x, \chi) = g_{\xi_1}(x) G_{\xi_2|\xi_1 = x}(\chi(x)), \tag{4.42}$$

where

$$G_{\xi_2|\xi_1 = x}(\chi(x)) = \Phi\left(\frac{\chi(x) - \mu_2 - \frac{\rho\sigma_2(x - \mu_1)}{\sigma_1}}{\sqrt{\sigma_2^2(1 - \rho^2)}}\right)$$

with $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{z^2}{2}} dz$.

Proof.

$$\begin{aligned}\frac{\partial\hat{\varphi}_1(x,\chi)}{\partial x} =& \frac{1}{2\pi\sqrt{|\Sigma|}}\int_{-\infty}^{\chi(x)} \exp\left(-\frac{1}{2}\Big(x-\mu_1, s-\mu_2\Big)\Sigma^{-1}\begin{pmatrix} x-\mu_1 \\ s-\mu_2 \end{pmatrix}\right) ds \\ =& g_{\xi_1}(x)\frac{1}{2\pi\sqrt{|\Sigma|}}\int_{-\infty}^{\chi(x)} \frac{\exp\left(-\frac{1}{2}\Big(x-\mu_1, s-\mu_2\Big)\Sigma^{-1}\begin{pmatrix} x-\mu_1 \\ s-\mu_2 \end{pmatrix}\right)}{g_{\xi_1}(x)} ds \\ =& g_{\xi_1}(x)\int_{-\infty}^{\chi(x)} \frac{g_{\xi}(x,s)ds}{g_{\xi_1}(x)} \\ =& g_{\xi_1}(x)\Phi\left(\frac{\chi(x)-\mu_2-\frac{\rho\sigma_2(x-\mu_1)}{\sigma_1}}{\sqrt{\sigma_2^2(1-\rho^2)}}\right).\end{aligned}$$

The last step follows from the standardization of the conditional probability distribution of a non-degenerate normal distribution, $\xi_2\,|\xi_1 \sim \mathcal{N}\left(\mu_2+\frac{\rho\sigma_2(x-\mu_1)}{\sigma_1}, (1-\rho^2)\sigma_2^2\right)$ (see, e.g., Theorem 7.3 [59]). □

4.6 The baker's Dynamic Optimization Model in $\mathcal{X}$

Consider a baker who has to decide about production quantity $x_1 \in \mathbb{R}_+$ of dough to satisfy the stochastic demand ξ_1 for bread in the morning ($t = 1$). The cost of production in the morning is $c_1 > 0$ per quantity of dough. At mid-day, $t = 2$, he takes into account the morning demand in order to decide how much dough $\chi_2(\xi_1) \in \mathbb{R}_+$ to produce to satisfy the stochastic demand ξ_2 for the afternoon. The cost of production in the afternoon is $c_2 > 0$ per quantity of dough. The baker would like to minimize his costs subject to satisfying the demand for bread with a high probability $p \in [0,1]$.
The baker's optimization problem is given by

$$\begin{aligned}\min_{(x_1,\chi_2)\in\mathbb{R}_+\times L^2(\mathcal{S}_1)} \quad & J(x_1,\chi_2) := c_1x_1 + \mathbb{E}[c_2\chi_2(\xi_1)] \\ \text{s.t.} \quad & \mathbb{P}\left(\xi_1 \le x_1, \xi_2 \le \chi_2(\xi_1)\right) \ge p \\ & x_1 \ge 0\,;\, \chi_2 \ge 0 \ a.e.\end{aligned} \tag{4.43}$$

where we assume that $\xi = (\xi_1,\xi_2) \sim \mathcal{N}(\mu,\Sigma)$ is a non-degenerate Gaussian random vector, and $g_\xi : \mathbb{R}^2 \to \mathbb{R}$ denotes the joint probability density function. Also, $g_{\xi_i} : \mathbb{R} \to \mathbb{R}$ denotes the marginal probability density function with respect to the random variable ξ_i.

Its deterministic equivalent optimization problem reads

$$\begin{aligned} \min_{(x_1,\varkappa_2)\in[0,\infty)\times L^2(\mathbb{R})} \quad & J(x_1,\varkappa_2) = c_1x_1 + c_2\int_0^\infty g_{\xi_1}(r)\varkappa_2(r)dr \\ \text{s.t.} \quad & \hat{\varphi}_1(x_1,\varkappa_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{\varkappa_2(r)} g_\xi(r,s)drds \geq p \\ & x_1 \geq 0, \varkappa_2 \geq 0. \end{aligned} \tag{4.44}$$

Note that we use $\hat{\varphi}_1$ from (4.40) and for the Gaussian random vector we have that $\mathcal{S}_1 = \mathbb{R}$.

Remark 4.33. Now, we would like to provide necessary optimality conditions for the Baker's optimization problem; that is, we would use the generalized KKT Theorem (see Thm. 4.11). In order to proceed in this way, it turns out to be useful to fix x_1. The reason as follows: recalling Theorem 4.30 and its Corollary 4.32, we notice that $\hat{\varphi}_1$ is continuously partially differentiable with respect to x_1, only if $\varkappa_2$ is continuous. However, we don't know *a priori* whether $\varkappa_2$ is continuous. In contrast, $\hat{\varphi}_1$ is continuously partially differentiable with respect to $\varkappa_2$ at any $(x_1,\varkappa_2) \in \mathbb{R}\times L^2(\mathbb{R})$. Thus, we would like to deduce the KKT conditions for a reduced version of problem 4.43, where we fix $\bar{x}_1 \geq 0$.

We first prove that the objective function of the Baker's model is continuously partially differentiable on $\mathbb{R}\times L^2(\mathbb{R})$.

Lemma 4.34. *Assume that g_{ξ_1} is continuous. Then the objective function in model* (4.44) *has the following partial derivatives with respect to x_1 and with respect to $\varkappa_2$.*

$$\frac{\partial J}{\partial x_1}(x_1,\varkappa_2) = c_1 \tag{4.45}$$

$$\nabla_{\varkappa_2} J(x_1,\varkappa_2)(r) = c_2 g_{\xi_1}(r)\mathbf{1}_{[0,\infty)}(r). \tag{4.46}$$

Moreover, the partial derivatives are continuous over $\mathbb{R}\times L^2(\mathbb{R})$.

Proof. Firstly, we notice that the function J is linear and continuous in x_1 and $\varkappa_2$, then this function is differentiable and the derivatives are given by

$$\frac{\partial J}{\partial x_1}(x_1) = c_1 \tag{4.47}$$

$$\nabla_{\varkappa_2} J(x_1,\varkappa_2)(r) = \frac{\partial\langle c_2 g_{\xi_1}(r)\mathbf{1}_{[0,\infty)}(r),\varkappa_2(r)\rangle}{\partial\varkappa_2} = c_2 g_{\xi_1}(r)\mathbf{1}_{[0,\infty)}(r), \tag{4.48}$$

where $c_2 g_{\xi_1}(r)\mathbf{1}_{[0,\infty)}(r)$ is an element of the dual space of $L^2(\mathbb{R})$. □

In addition, we prove that the Robinson constraint qualification (RCQ) for our infinite-dimensional problem holds.

Lemma 4.35. *Consider the optimization problem* (4.43) *(with $\bar{x}_1 \geq 0$ fixed)*

$$\begin{aligned} \min \quad & J(\bar{x}_1, x_2) \\ s.t. \quad & -\hat{\varphi}_1(\bar{x}_1, x_2) + p \leq 0 \\ & x_2 \in L^2(\mathbb{R}), \\ & x_2 \geq 0 \ a.e. \end{aligned} \tag{4.49}$$

Let $x_2^* \in L^2(\mathbb{R})$ be a feasible solution to problem (4.49)*; then it satisfies the Robinson Constraint Qualification.*

Proof. According to Definition 4.10, in order to prove that RCQ holds true for the baker's constraint set (4.43), we first write it in the form of (4.2), where $Z = L^2(\mathbb{R})$, $Q = L^2_+(\mathbb{R})$, $K = \mathbb{R}_+$ and $G(z) = \hat{\varphi}_1(z) - p$ with $z = (\bar{x}_1, x_2)$. Let $x_2^* \in L^2(\mathbb{R})$ be a feasible solution to problem (4.49). We know by Corollary 4.32 that $\nabla_{x_2}\hat{\varphi}_1(\bar{x}_1, x_2^*) - p$ is continuous in x_2^*. In this setting, we claim that Robinson's Constraint Qualification holds at x_2^*, i.e., we claim that

$$0 \in \operatorname{int}\{\hat{\varphi}_1(\bar{x}_1, x_2^*) - p + \nabla_{x_2}\hat{\varphi}_1(\bar{x}_1, x_2^*)(L^2_+(\mathbb{R}) - x_2^*) - \mathbb{R}_+\}.$$

We first observe that (by definition of Minkowski sum)

$$\begin{aligned} &\operatorname{int}\left\{\hat{\varphi}_1(\bar{x}_1, x_2^*) - p + \nabla_{x_2}\hat{\varphi}_1(\bar{x}_1, x_2^*)\big(L^2_+(\mathbb{R}) - x_2^*\big) - \mathbb{R}_+\right\} \\ &= \bigcup_{y \in L^2_+(\mathbb{R})} \operatorname{int}\left\{\hat{\varphi}_1(\bar{x}_1, x_2^*) - p + \langle \nabla_{x_2}\hat{\varphi}_1(\bar{x}_1, x_2^*), y - x_2^*\rangle - \mathbb{R}_+\right\}. \end{aligned}$$

Hence, it is sufficient to show that there exists $y \in L^2_+(\mathbb{R})$ such that

$$0 \in \operatorname{int}\{\hat{\varphi}_1(\bar{x}_1, x_2^*) - p + \langle \nabla_{x_2}\hat{\varphi}_1(\bar{x}_1, x_2^*), y - x^*\rangle - \mathbb{R}_+\}.$$

Assume $\hat{\varphi}_1(\bar{x}_1, x_2^*) - p =: \beta \geq 0$ and let $y(r) := x_2^*(r) + \mathbf{1}_{[-1,0]}(r) \geq 0$ where $y \in L^2_+(\mathbb{R})$. Firstly, note that by (4.41) it follows that

$$\nabla_{x_2}\hat{\varphi}_1(\bar{x}_1, x_2^*)(r) = g(r, x_2^*(r))\mathbf{1}_{(-\infty, \bar{x}_1]}(r),$$

so

$$\langle \nabla_{x_2} \hat{\varphi}_1(\bar{x}_1, x_2^*), y - x_2^* \rangle = \langle g(r, x_2^*) \mathbf{1}_{(-\infty, \bar{x}_1]}(r), \mathbf{1}_{[-1,0]}(r) \rangle = \int_{-1}^{0} g(r, x^*) dr =: \alpha > 0$$

since Gaussian density is strictly positive. Then

$$0 \in \text{int}\{\beta + \alpha - \mathbb{R}_+\} = int\{(-\infty, \beta + \alpha]\}.$$

□

Theorem 4.36. *Fix $\bar{x}_1 \in \mathbb{R}_+$ and let $x_2^* \in L^2(\mathbb{R})$ be a local solution to problem* (4.49)*. Assume that $(\xi_1, \xi_2) \sim \mathcal{N}(\mu, \Sigma)$ is a non-degenerate Gaussian random vector with*

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

and $|\rho| \neq 1$. Then there exists a non-negative continuous function $h \in L^2(\mathbb{R})$ and $\eta \in \mathbb{R}$ such that the optimal decision rule x_2^* satisfies

$$x_2^*(r) = \begin{cases} mr + \eta & \text{if } h(r) = 0 \\ 0 & \text{if } h(r) > 0 \end{cases}$$

with $r \in (-\infty, \bar{x}_1]$ and where the slope is given by $m := \frac{\sigma_2 \rho}{\sigma_1}$.

Proof. Fix $\bar{x}_1 \geq 0$ and let $x_2^* \in L^2(\mathbb{R})$ be a feasible solution to (4.43). We know by Lemma 4.34 and Theorem 4.30 that J and φ have continuous partial derivatives with respect to the second argument. Moreover, by Lemma 4.35 we have that Robinson's CQ holds at x_2^*. Therefore, by [21, Thm. 3.9] there exist Lagrangian multipliers $\lambda_1 \geq 0$ and $\lambda_2 : \mathbb{R} \to \mathbb{R}_+$ with $\lambda_2 \in L^2(\mathbb{R})$ such that

$$\nabla_{x_2} J(\bar{x}_1, x_2^*)(r) - \lambda_1 \nabla_{x_2} \varphi_1(x_1, x_2^*)(r) = \lambda_2(r) \text{ for almost all } r \tag{4.50}$$

$$p - \varphi_1(x_1, x_2^*) \leq 0 \tag{4.51}$$

$$\lambda_1 \left(p - \varphi_1(x_1, x_2^*)\right) = 0, \tag{4.52}$$

$$-x_2^* \leq 0 \tag{4.53}$$

$$\lambda_2(r)(-x_2^*(r)) = 0 \text{ for almost all } r. \tag{4.54}$$

From (4.50) and (4.54) we have that

$$x_2^*(r) \left(\nabla_{x_2} J(\bar{x}_1, x_2^*)(r) - \lambda_1 \nabla_{x_2} \varphi_1(x_1, x_2^*)(r)\right) = 0. \tag{4.55}$$

Thus, either $\varkappa_2^*(r) = 0$ or $\nabla_{\varkappa_2} J(\bar{x}_1, \varkappa_2^*)(r) - \lambda_1 \nabla_{\varkappa_2} \varphi_1(x_1, \varkappa_2^*)(r) = 0$. Assume the latter holds and let $r \in (-\infty, x_1]$. We have from (4.50), (4.46) and (4.41) that

$$c_2 g_{\xi_1}(r) = \lambda_1 g_{\xi}(r, \varkappa_2^*(r)) \tag{4.56}$$

$$\frac{c_2}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}(\frac{r-\mu_1}{\sigma_1})^2} = \frac{\lambda_1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[(\frac{r-\mu_1}{\sigma_1})^2 - 2\rho\frac{(r-\mu_1)(\varkappa_2(r)-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{(\varkappa_2(r)-\mu_2)}{\sigma_2}\right)^2\right]} \tag{4.57}$$

Let $\kappa_1 = \ln\left(\frac{c_2}{\sqrt{2\pi}\sigma_1}\right)$ and $\kappa_2 = \ln\left(\frac{\lambda_1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\right)$ with $|\rho| \neq 1$. It follows from (4.57)

$$\begin{aligned}
\kappa_1 - \frac{1}{2}(\frac{r-\mu_1}{\sigma_1})^2 &= \kappa_2 - \frac{1}{2(1-\rho^2)}\left[\left(\frac{r-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(r-\mu_1)(\varkappa_2(r)-\mu_2)}{\sigma_1\sigma_2}\right. \\
&\quad \left. + \left(\frac{\varkappa_2(r)-\mu_2}{\sigma_2}\right)^2\right] \\
\kappa_1 - \kappa_2 &= -\frac{1}{2(1-\rho^2)}\left[\frac{\rho^2(r-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(r-\mu_1)(\varkappa_2(r)-\mu_2)}{\sigma_1\sigma_2}\right. \\
&\quad \left. + \left(\frac{\varkappa_2(r)-\mu_2}{\sigma_2}\right)^2\right] \\
-2(1-\rho^2)(\kappa_1-\kappa_2) &= \left[\left(\frac{\varkappa_2(r)-\mu_2}{\sigma_2}\right)^2 - 2\rho\frac{(r-\mu_1)(\varkappa_2(r)-\mu_2)}{\sigma_1\sigma_2} + \right. \\
&\quad \left. \frac{\rho^2(r-\mu_1)^2}{\sigma_1^2}\right]
\end{aligned} \tag{4.58}$$

Let $z(r) = \frac{\varkappa_2(r)-\mu_2}{\sigma_2}$. It follows from (4.58) that

$$z(r)^2 - 2\rho\frac{(r-\mu_1)}{\sigma_1}z(r) + 2(1-\rho^2)(\kappa_1-\kappa_2) + \frac{\rho^2(r-\mu_1)^2}{\sigma_1^2} = 0. \tag{4.59}$$

That is

$$\begin{aligned}
z(r) &= 2\rho\frac{(r-\mu_1)}{2\sigma_1} \pm \frac{1}{2}\sqrt{\left(\frac{2\rho(r-\mu_1)}{\sigma_1}\right)^2 - 4\left(\frac{2\sigma_1^2(\kappa_1-\kappa_2)(1-\rho^2) + \rho^2(r-\mu_1)^2}{\sigma_1^2}\right)} \\
&= \frac{\rho(r-\mu_1)}{\sigma_1} \pm \frac{1}{2}\sqrt{-8(\kappa_1-\kappa_2)(1-\rho^2)}.
\end{aligned}$$

Substituting $\varkappa_2(r)$ in the previous equation, we have

$$\varkappa_2(r) = \frac{\sigma_2\rho}{\sigma_1}r \pm \frac{\sigma_2}{2}\sqrt{-8(\kappa_1-\kappa_2)(1-\rho^2)} - \frac{\sigma_2\rho\mu_1}{\sigma_1} + \mu_2. \tag{4.60}$$

Let $m = \frac{\sigma_2 \rho}{\sigma_1}$ and $\eta = \pm \frac{\sigma_2}{2} \sqrt{-8(\kappa_1 - \kappa_2)(1-\rho^2)} - \frac{\sigma_2 \rho \mu_1}{\sigma_1} + \mu_2$, then by (4.55) we have that

$$x_2^*(r) = \begin{cases} mr + \eta & \text{if } \lambda_2(r) = 0, \\ 0 & \text{if } \lambda_2(r) > 0. \end{cases} \tag{4.61}$$

□

In most of the models of mathematical optimization in economics, non-negativity conditions are required. But sometimes it might be useful to consider the non-negativity conditions separately or exclude them completly. If we consider problem (4.49) without restricting $x_2 \geq 0$, the optimal decision policy is an affine linear solution.

Corollary 4.37. *Consider a simple version of the optimization problem* (4.43) *with* $\bar{x_1} \geq 0$ *fixed*

$$\begin{aligned} \min_{(\bar{x_1}, x_2) \in \mathbb{R} \times L^2(\mathbb{R})} \quad & J(x_1, x_2) \\ s.t. \quad & -\hat{\varphi}_1(x_1, x_2) + p \leq 0 \\ & x_2 \in L^2(\mathbb{R}), \end{aligned} \tag{4.62}$$

and let $x_2^* \in L^2(\mathbb{R})$ *be a solution. Assume that* $\xi \sim \mathcal{N}(\mu, \Sigma)$ *is a non-degenerate Gaussian random vector. Then* x_2^* *is an affine linear solution. Moreover, our infinite-dimensional problem under dynamic joint probabilistic constraints* (4.62) *can be reduced to a numerically tractable finite-dimensional static probabilistic constrained problem of the type* (1.27).

Proof. Let $x_2^* \in L^2(\mathbb{R})$ be a solution to problem (4.62). From Theorem 4.36, in particular from equation (4.61), we have that $\lambda_2(r) \equiv 0$ (since we do not have non-negativity condition for x_2) so, $x_2^* = mr + \eta$ holds with $m = \rho \frac{\sigma_2}{\sigma_1}$. In this setting, our infinite-dimensional problem can be reduced to a two-dimensional static probabilistic constrained problem:

$$\begin{aligned} \min_{(x_1, \eta) \in \mathbb{R}^2} \quad & c_1 x_1 + c_2 \int_0^\infty g_{\xi_1}(r)\,(mr + \eta)\,dr \\ \text{s.t.} \quad & \mathbb{P}\left(T\xi \leq \begin{bmatrix} x_1 \\ \eta \end{bmatrix} \right) \geq p, \end{aligned} \tag{4.63}$$

where $T = \begin{pmatrix} 1 & 0 \\ -m & 1 \end{pmatrix}$ and $\nu := T\xi \sim \mathcal{N}(T\mu, T\Sigma T^T)$ with

$$T\Sigma T^T = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2(1-\rho) \end{pmatrix}.$$

□

Corollary 4.38. *Let* $(x_1^*, \varkappa_2^*) \in \mathbb{R} \times L^2(\mathbb{R})$ *be a solution to problem* (4.62). *Assume that* $\xi \sim \mathcal{N}(\mu, \Sigma)$ *is a non-degenerate Gaussian random vector and let* $r \leq x_1^*$. *Then the following statements are satisfied:*

1. $\varkappa_2^*(r)$ *is a static decision rule if and only if* $\rho = 0$.
2. *If* $\rho > 0$, *then* $\varkappa_2^*(r)$ *has a positive slope.*
3. *If* $\rho < 0$, *then* $\varkappa_2^*(r)$ *has a negative slope.*
4. *If* $\rho \to \pm 1$, *then the slope* $m \to \pm \frac{\sigma_2}{\sigma_1}$.

Proof. The results follow from Theorem 4.36, Corollary 4.37 and the fact that standard deviation is positive. □

4.7 Two-stage single hydro reservoir model in $\mathcal{X}$

In this section, we consider a two-period model of a single water reservoir for hydroelectricity operation model with uncertain water inflow (4.7). We may think of the two-stage model as consisting of a summer $(t = 1)$ and a winter period $(t = 2)$, where there is uncertainty about the water inflow ξ_1, ξ_2 in both periods. We assume that there is a one-to-one correspondence between water released in m^3 and electricity produced measured in kWh. The key economic question in hydro-power management is: how much energy x_1 should I convert at the beginning of summer, without knowing how much precipitation will occur at the end of the summer ξ_1, and how much water l_1 should be stored at the end of summer. Then, by considering the information of the past, how much water $\varkappa_2(\xi_1)$ should one release in winter with the uncertainty of how much water inflow ξ_2 will occur during winter. The sequence of decisions is

$$\boxed{x_1 \curvearrowright \xi_1 \curvearrowright \varkappa_2(\xi_1) \curvearrowright \xi_2.}$$

The operation of hydro-power is thus a dynamic one. Taking the information into account, due to the successive realization of random inflows, we denote our release policy as $(x_1, \varkappa_2(\xi_1))$. Moreover, the technical and environmental constraints (4.5) for the two-stage dynamic probabilistic constrained function can be explicitly given by

$$\varphi_3(x, \chi) := \mathbb{P}(l_* \leq l_0 + \xi_1 - x \leq l^*, l_* \leq l_0 + \xi_1 + \xi_2 - x + \chi(\xi_1) \leq l^*), \qquad (4.64)$$

with constants $a = c = l_* - l_0$ and $b = d = l^* - l_0$. The next corollary presents the explicit partial derivatives for the case of the probability function in the hydro-power reservoir (4.64), where we assume that $\xi \sim \mathcal{N}_2(\mu, \Sigma)$.

Corollary 4.39. *Consider the probability function defined in* (4.64)*, where ξ is a non-degenerate Gaussian random vector with mean $\mu = (\mu_1, \mu_2)$ and covariance matrix Σ. Then the probability function* (4.64) *is (continuously) partially differentiable with respect to $\varkappa$ at any $(x, \varkappa) \in \mathbb{R} \times L^2(\mathbb{R})$, and the following formula holds*

$$\begin{aligned}\nabla_\varkappa \varphi_3(x,\varkappa)(u) &= \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\Bigg(-\frac{1}{2}\begin{pmatrix} u-\mu_1 \\ u + l^* - l_0 + x + \varkappa(u) - \mu_1 - \mu_2 \end{pmatrix}^\top \Sigma^{-1} \\ &\quad \begin{pmatrix} u-\mu_1 \\ u + l^* - l_0 + x + \varkappa(u) - \mu_1 - \mu_2 \end{pmatrix}\Bigg) \mathbf{1}_{[\bar{l}_* - l_0 + x, l^* - l_0 + x]}(u) \\ &\quad - \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\Bigg(-\frac{1}{2}\begin{pmatrix} u-\mu_1 \\ u + l_* - l_0 + x + \varkappa(u) - \mu_1 - \mu_2 \end{pmatrix}^\top \Sigma^{-1} \\ &\quad \begin{pmatrix} u-\mu_1 \\ u + l_* - l_0 + x + \varkappa(u) - \mu_1 - \mu_2 \end{pmatrix}\Bigg) \mathbf{1}_{[l_* - l_0 + x, l^* - l_0 + x]}(u). \quad (4.65)\end{aligned}$$

In addition, if $\varkappa$ is continuous, then the probability function (4.64) *is also partially differentiable with respect to x, and its partial derivative is given by*

$$\begin{aligned}\frac{\partial \varphi_3(x,\varkappa)}{\partial x} &= \frac{1}{2\pi\sqrt{|\Sigma|}} \int_{l_* - l_0 + x + \varkappa(x)}^{l^* - l_0 + x + \varkappa(x)} \exp\Bigg(-\frac{1}{2}\Big(x_1 - \mu_1, x_1 + s - \mu_1 - \mu_2\Big)\Sigma^{-1} \\ &\quad \begin{pmatrix} x_1 - \mu_1 \\ x_1 + s - \mu_1 - \mu_2 \end{pmatrix}\Bigg) ds.\end{aligned}$$

Proof. First, let us notice that considering the random vector $\tilde{\xi} = (\xi_1, \xi_1 + \xi_2)$, we can write the probability function (4.64) in the form

$$\mathbb{P}(l_* \le l_0 + \xi_1 - x \le l^*, l_* \le l_0 + \tilde{\xi}_2 - x + \varkappa(\tilde{\xi}_1) \le l^*), \quad (4.66)$$

which coincides with the probability function (4.18). Consequently, by Observation 4.5.1 we notice that the random vector $\tilde{\xi}$ has covariance matrix

$$\tilde{\Sigma}^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \Sigma^{-1} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

and mean $\tilde{\mu} = (\mu_1, \mu_1 + \mu_2)$. Then by applying Corollary 4.31 with the constants $a = c = l_* - l_0$ and $b = d = l^* - l_0$ we get the result.

□

In order to study the two-stage single hydro reservoir model, an objective function has to be specified. Cost minimization or maximization of social welfare are usually applied by social planners or policy makers in the energy sector as opposed to the profit maximization of private investors [120, ch.10] and [79]. Theoretically, when every private company tries to maximize its benefit, both central planning and private perspective should lead to the same outcome under perfect market competition and without transmission constraints. However, real markets are seldom perfect. Nonetheless, the existence of market distortions, i.e. market power, together with the deregulation of the energy market, do not mean that central planning has become obsolete. On the contrary, the result of the maximization of welfare is commonly used as a benchmark for measuring the health of the system and for identifying the required corrective actions to be accomplished by policy makers [67]. Thus, we will consider two objective functions for the same probabilistic constraint (4.64). For simplicity of the results, in the following, we set $\mathcal{X} = \mathbb{R} \times L^2(\mathcal{S}_1)$.

4.7.1 Model for a producer

Firstly, we analyze the problem for the producer, which is to maximize the water release in both stages, without emptying or flooding the reservoir. In this case, we have a linear functional $J_1 : \mathcal{X} \to \mathbb{R}$ and the model is represented by

$$\begin{aligned} \max_{(x_1, \varkappa_2) \in \mathbb{R} \times L^2(\mathbb{R})} \quad & J_1[x_1, \varkappa_2] := p_1 x_1 + p_2 \mathbb{E}\left(\varkappa_2(\xi_1)\right) \\ \text{s.t.} \quad & \mathbb{P}(l_* \leq l_0 + \xi_1 - x \leq l^*, l_* \leq l_0 + \xi_1 + \xi_2 - x + \varkappa(\xi_1) \leq l^*) \geq p \\ & x_1, \varkappa_2(\xi_1) \geq 0 \quad \mathbb{P}\text{-almost surely.} \end{aligned} \tag{4.67}$$

The released water is used to sell hydro-electricity at prices (p_1, p_2), which are assumed to be known in advance.

Numerical results for the producer problem

Here we illustrate the solutions to the two-stage hydro-power model for the producer (4.67). For the subsequent numerical results, we employed a gradient-based framework SQP (profiting from (4.65)) in MATLAB. For this model, we consider the water reservoir problem with the following data:

$$l_* = 1, l^* = 16, l_0 = 10, p = 0.9, p_1 = 30 = p_2$$

$$\xi \sim \mathcal{N}\left((10,10), \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}\right),$$

where the latter expression means that the random precipitation $\xi = (\xi_1, \xi_2)$ has a bivariate normal distribution with independent components, each of them having mean 10 and variance 4 (or standard deviation 2).

First, we approximate $\chi_2(r)$ by a piecewise linear function $\hat{\chi}_2(r)$. Let $I = [a, b]$ be a given interval. We partition I by a mesh with $[c_i, c_{i+1}]$ such that

$$\hat{\chi}_2(r) := \begin{cases} \beta_1 + s_1(r - c_1) & c_1 \leq r \leq c_2 \\ \beta_2 + s_2(r - c_2) & c_2 \leq r \leq c_3 \\ \vdots & \\ \beta_n + s_n(r - c_n) & c_n \leq r \leq c_{n+1}. \end{cases}$$

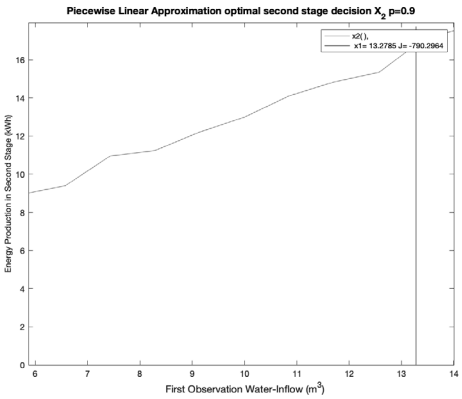
The producer's problem (4.67) can be approximated by

$$\begin{aligned} \min_{(x_1, s_1, \ldots, s_n, \beta_1, \ldots, \beta_n) \in \mathbb{R}^{2n+1}} \quad & -p_1 x_1 - p_2 \int_{x_1 - l_0 + l_*}^{x_1 - l_0 + l^*} \hat{\chi}_2(r) g_{\xi_1}(r) dr \\ \text{s.t.} \quad & p - \int_{\underline{\phi}_1(x_1)}^{\overline{\phi}_1(x_1)} \int_{\underline{\tilde{\phi}}_2(x_1, s_1, \ldots, s_n, \beta_1, \ldots, \beta_n)(r)}^{\tilde{\overline{\phi}}_2(x_1, s_1, \ldots, s_n, \beta_1, \ldots, \beta_n)(r)} g_\xi(r, s) ds dr \leq 0 \\ & x_1, \hat{\chi}_2 \geq 0 \quad \mathbb{P}\text{-almost surely,} \end{aligned} \tag{4.68}$$

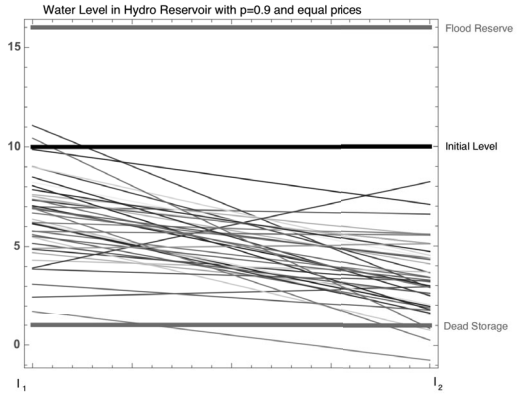
where $g_\xi(r, s)$ is the joint Gaussian density and $g_{\xi_1}(r)$ is the marginal density and the functions for the limits of the integrals are

$$\begin{aligned} \overline{\phi}_1(x_1) &= l^* - l_0 + x_1, \\ \underline{\phi}_1(x_1) &= l_* - l_0 + x_1, \\ \hat{\tilde{\phi}}_2(x_1, s_1, \ldots, s_n, m_1, \ldots, m_n)(r) &= l^* - l_0 + x_1 + \hat{\chi}_2(r) - r, \\ \hat{\tilde{\phi}}_2(x_1, s_1, \ldots, s_n, m_1, \ldots, m_n)(r) &= l_* - l_0 + x_1 + \hat{\chi}_2(r) - r. \end{aligned} \tag{4.69}$$

Figure 4.3a illustrates the optimal decision policies x_1^* and χ_2^* for the two-stage hydro-power problem under the linear objective function of the producer. In this case, χ_2^* corresponds to an affine linear function of the first-stage water inflow. With a probability level $p = 0.9$ Figure 4.3b illustrates the corresponding water level l_1 in th first-stage (at the left side of the graph) and the corresponding level of water l_2 in

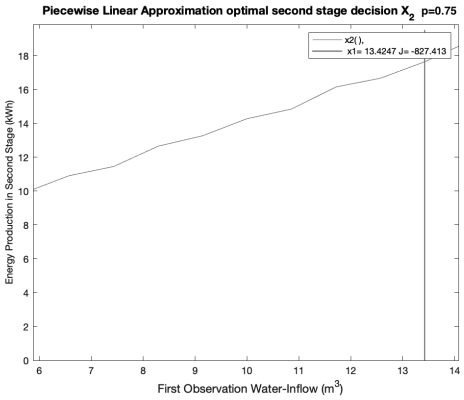


(A) Illustration of optimal second stage decision for the producer with $p = 0.9$ and 11 pieces.

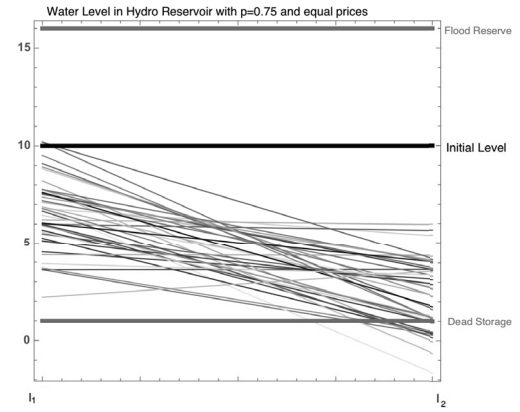


(B) Illustration of the corresponding water level with 40 random inflow scenarios, from which 4/40 are not within the given (red) limits.

FIGURE 4.3



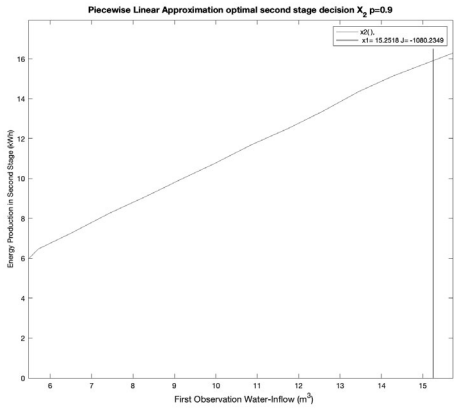
(A) Illustration of second stage decision for the Producer with lower probability $p = 0.75$ and 11 pieces.



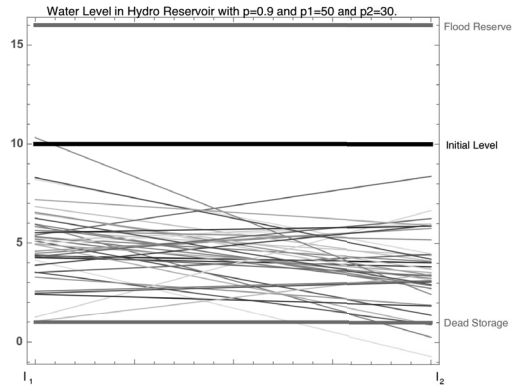
(B) Illustration of the corresponding water level with 40 random inflow scenarios, from which 10/40 are not within the given limits.

FIGURE 4.4

the second-stage (at the right side). Exactly ten percent of the scenarios are outside the given limits. Figure 4.4a illustrates the optimal decision policy x_2^* when we lower the probability level to $p = 0.75$, in this case, we get a higher profit $J_1^* = 827$ than in Figure 4.3a. However, as figure 4.4b depicts, twenty five percent of the scenarios are out of the given dead storage in l_2. Lastly, Figure 4.5a and 4.6a depict the optimal decision policy x_2^* when prices are different. When p_1 is higher, the optimal release of the first stage x_1^* is much higher than in Figure 4.3a. Likewise, when p_2 is higher, as in Figure 4.6a, the optimal release of the second stage x_2^* shifts upwards.

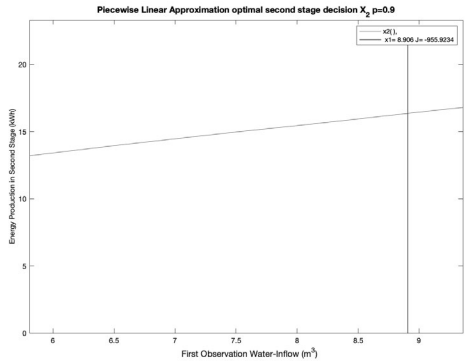


(A) Illustration of second stage decision for the Producer with different prices $p_1 = 50$ $p_2 = 30$, $p = 0.9$ and 11 pieces

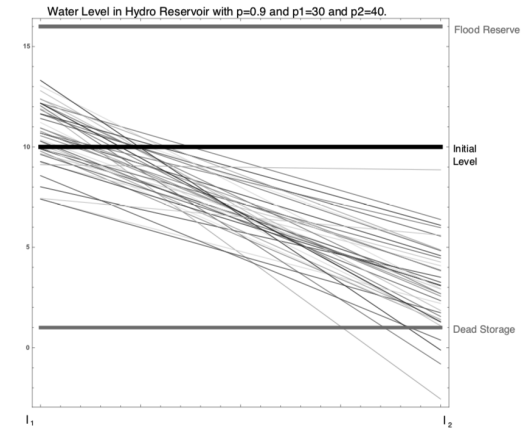


(B) Illustration of the corresponding water level with 40 random inflow scenarios, from which 4/40 are not within the given limits.

FIGURE 4.5



(A) Illustration of second stage decision for the Producer with different prices $p_1 = 30$ and $p_2 = 40$, $p = 0.9$ and 11 pieces.



(B) Illustration of the corresponding water level with 40 random inflow scenarios, from which 4/40 are not within the given (red) limits.

FIGURE 4.6

4.7.2 Model for a social planner

Secondly, we analyze the problem of a social planner or policy maker. In the case of hydro-power, water inflow (rainfall or snow melting) is the only variable cost; thus there are zero operating costs (marginal costs). Therefore, the minimizing costs approach cannot be applied. This is why we focus on the maximization of the total social welfare. The objective function $J_2 : \mathcal{X} \to \mathbb{R}$ describes the social welfare. We recall that the social welfare is defined as the sum of the consumer and producer surplus. Consumer surplus is the difference between what consumers are willing to pay and what they actually pay. Likewise, producer surplus is the difference

between what the producer gets paid and the marginal costs of production. Since the marginal costs are zero in hydro-power management, the social welfare is simplified to the area under the inverse consumer demand function $\rho_k(\varkappa_k)$ for $k = 1, 2$. It is further assumed that the inverse demand ρ_k is right-continuous and non-increasing for all periods. The model for the social planner is given by

$$\begin{aligned} \max_{(x_1,\varkappa_2)\in\mathbb{R}\times L^2(\mathbb{R})} \quad & J_2[x_1,\varkappa_2] := \int_0^{x_1}\rho_1(z)dz + \mathbb{E}\left[\int_0^{\varkappa_2(\xi_1)}\rho_2(z)dz\right] \\ \text{s.t.} \quad & \mathbb{P}(l_* \le l_0 + \xi_1 - x \le l^*, l_* \le l_0 + \xi_1 + \xi_2 - x + \varkappa(\xi_1) \le l^*) \ge p \\ & x_1, \varkappa_2(\xi_1) \ge 0 \ \ \mathbb{P}\text{-almost surely.} \end{aligned} \tag{4.70}$$

Both functions J_1 and J_2 are standard objective mappings in hydro-power economics (see [67], [51, ch. 3]).

Properties of the Objective Function J_2

For convenience, from now on, we will assume that

(A_1) $\mathcal{S}_1$ is an open interval, and $0 < i_1 := \inf_{x\in\mathcal{S}_1} g_{\xi_1}(x)$ and $s_1 = \sup_{x\in\mathcal{S}_1} g_{\xi_1}(x) < \infty$.[1]

(A_2) For $j = 1, 2$ the prices are linear $\rho_j(z) = m_j z + n_j$ and $m_j < 0$ (satisfies the economic law of diminishing demand).

Under the above assumptions, we have that the objective function of (4.70) is

$$J_2[x_1,\varkappa_2] = \frac{m_1}{2}x_1^2 + n_1x_1 + \frac{m_2}{2}\mathbb{E}(\varkappa_2^2(\xi)) + n_2\mathbb{E}(\varkappa_2(\xi)). \tag{4.71}$$

Next, we prove that under our assumptions (A_1) and (A_2), J_2 is a concave map and satisfies some coercive properties in $\mathcal{X}$.

Proposition 4.40. *The function J_2 defined in* (4.70) *is concave, continuous over $\mathcal{X}$, and satisfies that*

$$\lim_{\|\varkappa\|\to\infty} J_2(\varkappa) = -\infty. \tag{4.72}$$

[1]Under our assumption (A_1), the random vector ξ, which represents the random inflow of water behaves like a truncated Gaussian distribution, since it is unrealistic to believe that for any amount of water, the probability of raining more than this amount is positive.

Proof. First we notice that J_2 can be rewritten as

$$J_2[x_1, \varkappa_2] = \frac{m_1}{2}x_1^2 + n_1 x_1 + \frac{m_2}{2}\mathbb{E}(\varkappa_2^2(\xi)) + n_2\mathbb{E}(\varkappa_2(\xi)) \tag{4.73}$$

$$= \frac{m_1}{2}x_1^2 + n_1 x_1 + \frac{m_2}{2}\int_{\mathcal{S}_1} \varkappa_2^2(x) g_{\xi_1}(x)dx + n_2 \int_{\mathcal{S}_1} \varkappa_2(x) g_{\xi_1}(x)dx. \tag{4.74}$$

Thus, (4.74) implies that J is concave due to the fact that $m_j < 0$. Moreover, it is easy to see that all the terms in (4.74) are continuous with the exception of the expression $\frac{m_2}{2}\int_{\mathcal{S}_1} \varkappa_2^2(x) g_{\xi_1}(x)dx$. However, we have the following inequality

$$\frac{m_2}{2}s_1 \int_{\mathcal{S}_1} \varkappa_2^2(x)dx, \leq \frac{m_2}{2}\int_{\mathcal{S}_1} \varkappa_2^2(x) g_{\varkappa_1}(x)dx \leq \frac{m_2}{2}i_1 \int_{\mathcal{S}_1} \varkappa_2^2(x)dx, \tag{4.75}$$

where the terms i_1 and s_1 are defined in (A_1). This implies that J is bounded from below (by a continuous concave function), which yields the continuity of the entire function J_2. Now, using Cauchy-Schwarz's inequality, we obtain the inequality

$$n_2 \int_{\mathcal{S}_1} \varkappa_2(x) g_{\varkappa_1}(x)dx \leq |n_2|\lambda(\mathcal{S}_1)^{1/2} s_1^{1/2} \left(\int_{\mathcal{S}_1} \varkappa_2^2(x)dx\right)^{1/2}, \tag{4.76}$$

where $\lambda(\cdot)$ denotes the Lebesgue measure. Consequently, using (4.75) and (4.76) we get

$$J[x_1, \varkappa_2] \leq (\frac{m_1}{2}|x_1| + |n_1|)|x_1| + \left(\frac{m_2}{2}i_1 \left(\int_{\mathcal{S}_1} \varkappa_2^2(x)dx\right)^{1/2} + |n_2|s_1^{1/2}\right)\left(\int_{\mathcal{S}_1} \varkappa_2^2(x)dx\right)^{1/2}. \tag{4.77}$$

Now, we recall that $m_1 < 0$ and $m_2 < 0$, such that $\frac{m_1}{2}|x_1| + |n_1| < 0$ holds for large enough x_1. Similarly one can show that the inequality

$$\frac{m_2}{2}i_1 \left(\int_{\mathcal{S}_1} \varkappa_2^2(x)dx\right)^{1/2} + |n_2|s_1^{1/2} < 0,$$

if the norm of $\varkappa_2$ is large enough. Consequently, taking the limits in (4.77) we get (4.72). □

Remark 4.41. We would like to prove the existence of a solution to both models of producer and social planner of a hydro-power. To this end, we would like to use the direct method exposed in Proposition 4.4. Although both objective functions have

nice properties, this is not possible, since recalling from Theorem 4.17, our feasible set $\mathfrak{M}(p) := \{(x_1, \mathfrak{x}_2) \in \mathcal{X} : \varphi_3(x_1, \mathfrak{x}_2) \geq p\}$ is not weakly sequentially closed. Instead, taking advantage of the explicit partial derivatives of φ_3 given in Corollary 4.39, we provide numerical results for both models (4.67) and (4.70).

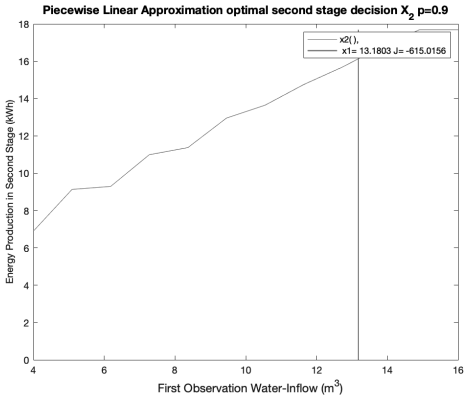
Numerical results for the social planner problem

Here, we illustrate the solutions to the two-stage hydro-power model for the social planner. For the subsequent numerical results, we have employed a gradient-based framework SQP (profiting from (4.65)) in MATLAB. We consider the water reservoir problem with the data:

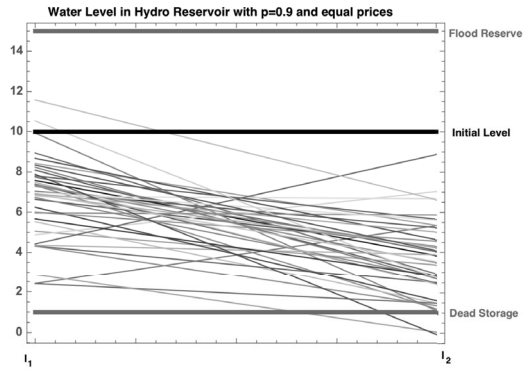
$$l_* = 1, l^* = 15, l_0 = 10, p = 0.9, m_1 = -1 = m_2, n_1 = 30 = n_2,$$
$$\xi \sim \mathcal{N}\left((10,10), \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}\right).$$

The social planner optimization problem (4.70) can be approximated by the finite-dimensional problem given by

$$\begin{aligned} \min_{(x_1,s_1,\ldots,s_n,\beta_1,\ldots,\beta_n)\in\mathbb{R}^{2n+1}} \quad & \frac{x_1^2}{2} - n_1 x_1 + \frac{1}{2}\int_{x_1-l_0+l_*}^{x_1-l_0+l^*} \hat{x_2}^2(r) q_{\xi_1}(r)dr - n_2 \int_{x_1-l_0+l_*}^{x_1-l_0+l^*} \hat{x_2}(r) g_{\xi_1}(r)dr \\ \text{s.t.} \quad & p - \int_{\underline{\phi}_1(x_1)}^{\overline{\phi}_1(x_1)} \int_{\underline{\tilde{\phi}}_2(x_1,s_1,\ldots,s_n,\beta_1,\ldots,\beta_n)(r)}^{\tilde{\overline{\phi}}_2(x_1,s_1,\ldots,s_n,\beta_1,\ldots,\beta_n)(r)} g_\xi(r,s)dsdr \leq 0 \\ & x_1, \hat{x_2} \geq 0 \quad \mathbb{P}-\text{almost surely.} \end{aligned} \tag{4.78}$$

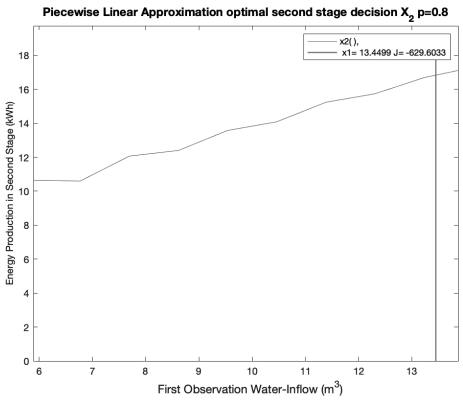


(A) Illustration of optimal second stage decision for social planner with $p = 0.9$ and equal prices.

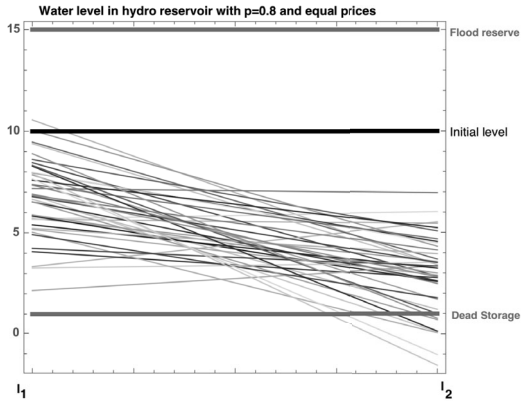


(B) Illustration of corresponding water level with 40 random inflow scenarios, from which 3/40 are not within the given dead storage limit.

FIGURE 4.7

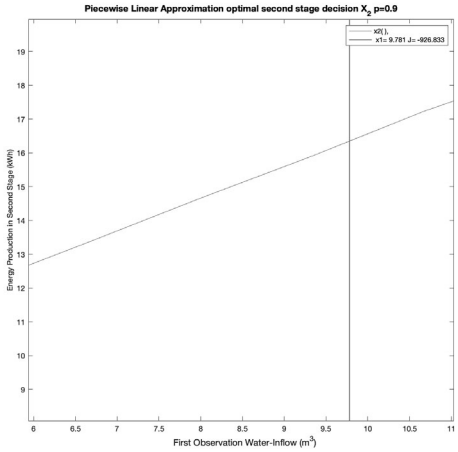


(A) Illustration of optimal second stage decision for social planner with $p = 0.8$ and equal prices.

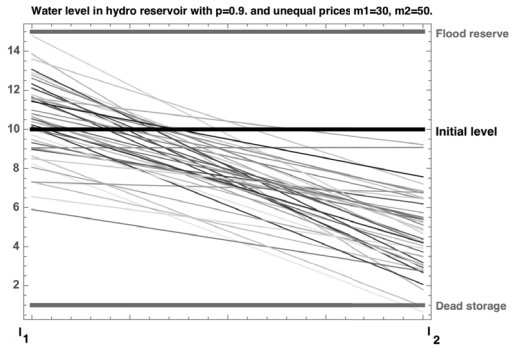


(B) llustration of corresponding water level with 40 random inflow scenarios, from which 8/40 are not within the given limits.

FIGURE 4.8



(A) Illustration of second stage decision with $p = 0.9$ and different price weights for each stage $m_1 = 30 < 50 = m_2$.



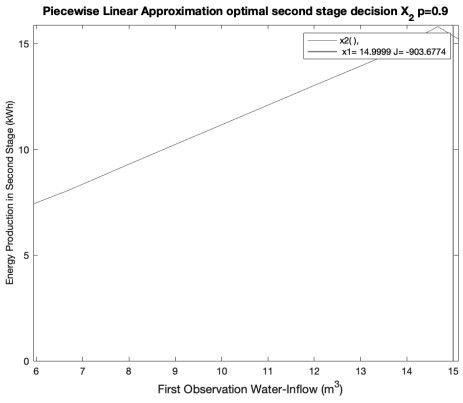
(B) Illustration of the water level with 40 random inflow scenarios, from which 2/40 are not within the given limits (in red) in second stage.

FIGURE 4.9

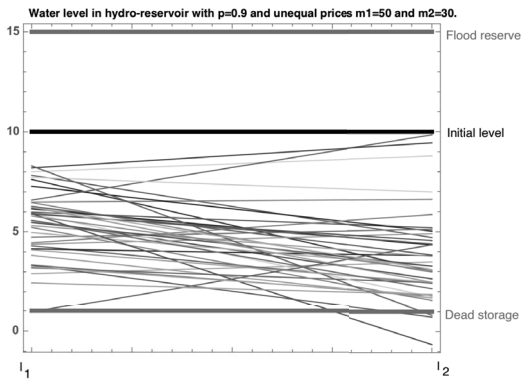
Figure 4.7a-4.10a illustrate the optimal decision policies of the first x_1^* and second stage x_2^* for the two-stage hydro-power problem under the nonlinear objective function of the social planner. For different indicated parameters, x_2^* corresponds to an affine linear function, similarly to the solution of the one-side Baker's model (4.62).

4.8 Two-stage single hydro reservoir model in $\mathcal{X}^1$

For the simplicity of our results, during this section, we set $\mathcal{X}^1 = \mathbb{R} \times W^{1,2}(\mathcal{S}_1)$. Consider the following optimization model for the social planner, where we add an



(A) Illustration of second stage decision with $p = 0.9$ and different price weights for each stage $m_1 = 50 > 30 = m_2$.



(B) Illustration of the water level with 40 random inflow scenarios, from which 3/40 are not within the given limits (in red) in second stage.

FIGURE 4.10

additional constraint to our optimization model (4.7). Now we consider a constraint of the variation of the control x_2 under change of the random vector ξ_1. Under this model with bounded variation on $\mathcal{X}^1$, from the mathematical point of view, we have transformed our optimization problem into a minimum norm control problem in a Sobolev space.

$$\begin{aligned} &\max_{(x_1,x_2)\in\mathcal{X}^1} J_2[x_1,x_2] \\ \text{s.t.}\quad &(x_1,x_2)\in\mathfrak{M}^1(p), \\ &x_1\geq 0, \text{ and } x_2\geq 0, \ \mathbb{P}\text{a.s.} \\ &x_2\in V(b), \end{aligned} \tag{P_b}$$

where $V(b) := \{x \in W^{1,2}(\mathcal{S}_1) : \int_{\mathcal{S}_1} |x'(s)|^2 ds \leq b\}$ is the set of policies with bounded variation.

4.8.1 Solution on $\mathcal{X}^1$

Let us establish the following lemma.

Lemma 4.42. *Consider the function $f : \mathcal{X}^1 \to \mathbb{R} \cup \{+\infty\}$ defined by*

$$f(x_1,x_2) = \begin{cases} -J_2[x_1,x_2] & \text{if } x_2\in V(b) \\ +\infty & \text{if } x_2\notin V(b). \end{cases} \tag{4.79}$$

Then f is convex, lower semi-continuous and satisfies

$$\lim_{\|x\|\to+\infty} f(x) = +\infty. \tag{4.80}$$

Proof. Using Proposition 4.40 and the fact that $V(b)$ is closed and convex, we can see that f is convex and lower semi-continuous over $\mathcal{X}^1$.

Now, let us prove (4.80). Consider a sequence $(x_{1,n}, \varkappa_{2,n}) \in \mathcal{X}^1$ such that

$$\|(x_{1,n}, \varkappa_{2,n})\|_{\mathcal{X}^1} \to +\infty,$$

then for large enough n, $\varkappa_{2,n} \in V(b)$ (otherwise $f(x_n) = \infty$ for an infinite number of n, and (4.80) holds trivially). The last implies that necessarily $\|x_n\|_{\mathcal{X}} \to +\infty$, and by Theorem 4.40 we have that (4.80) holds. □

Lemma 4.43. *The sets $V(b)$ and $\{\varkappa \in \mathcal{X}^1 : \varkappa \geq 0, \text{ a.s.}\}$ are weakly-closed.*

Proof. These sets are convex and closed with respect to the norm. Then, using the fact that $\mathcal{X}^1$ is a Banach space (actually a Hilbert space) and [29, Theorem 3.7] we have that these sets are weakly-closed. □

Theorem 4.44. *Let $b \geq 0$. If the problem* (P_b) *is feasible, then it admits a solution.*

Proof. Let us notice that we can write the problem (P_b) as a minimization problem of the following form

$$\min_C f, \tag{4.81}$$

where f is defined in (4.79) and C is the set of all $(x_1, \varkappa_2) \in \mathcal{X}^1$ such that $x_1, \varkappa(\xi_1) \geq 0$ $\mathbb{P}$-a.s. and $\varkappa \in \mathfrak{M}^1(p)$.

Now, consider a minimizing sequence $(x_{1,n}, \varkappa_{2,n}) \subset \mathcal{X}^1$ of the problem (4.81). Since the problem is feasible, we can assume that $\varkappa_{2,n} \in V(b) \cap C$, and by (4.80) we have that $(x_{1,n}, \varkappa_{2,n})$ must be bounded. Then, by the reflexivity of the space $\mathcal{X}^1$, there is a subsequence $(x_{1,n_k}, \varkappa_{2,n_k})$ that converges weakly to $(x_1, \varkappa_2)$. Then by Lemma 4.43, $x_1, \varkappa_2 \geq 0$ a.s. and $x_1, \varkappa_2 \in V(b)$. Moreover, by Proposition 4.13 we have that $\varkappa \in \mathfrak{M}^1(p)$. Finally, by Lemma 4.42 we have that $\min_C f \leq f(\varkappa) \leq \liminf f(\varkappa_n) = \min_C f$, concluding the proof. □

Remark 4.45. Our proof is analogous to the proof of Proposition 4.4, which we could have alternatively used. But we wanted to explicitly show two important points: Firstly, $\mathfrak{M}^1(p)$ is sequentially weakly continuous, and secondly, that $\varkappa_2 \in V(b)$ such that the norm $||\varkappa_2||_{W_{1,2}}$ is bounded.

Chapter 5

Conclusion

The main contributions of the thesis address the theoretical analysis, modeling and numerical treatment of two major novel topics in connection with probabilistic constraints: the recently defined family of *probust constraints* and *dynamic probabilistic constraints*.

The first two chapters offer an overview of classical approaches to solving optimization problems under uncertainty. We put substantial focus to probabilistic constraints. In chapter 2 we examine two algorithms for dealing with optimization problems under joint (linear and nonlinear) probabilistic constraints. For nonlinear probabilistic constraints with a Gaussian or Gaussian-like underlying stochastic vector, we exploit the spherical radial decomposition. Although this decomposition is widely used and known, for the sake of completeness, we provide our own proof. Moreover, in [140] a gradient formula is derived for a non-linear probability function, while in [141] (sub)-gradient formulae are derived for the case of multiple constraints with a fixed index set.

In chapter 3 we introduce probust functions for optimization problems having stochastic and non-stochastic uncertain parameters. A convenient combination with ideas from robust optimization then leads to probust constraints, i.e., probability functions acting on generalized semi-infinite inequality systems.

Moreover, we model and solve two problems from the viewpoint of a gas network operator. The methodology used in both problems is to first solve the robust counterpart of the inner robust problem of (3.5). We first obtain an analytical worst-case solution, which allows us to turn the infinite inequalities system of the probust constraint into a traditional joint probabilistic constraint (with a finite number of inequalities). In this way, we can solve the problem algorithmically using the spheric-radial decomposition.

In both of the applications, we assumed that we work with tree-structured gas networks, which helped us to analytically identify the inner worst-case solution. In

the presence of cycles, there is an implicit relation for the feasible loads and one cannot identify the worst-case solution analytically. In general, one cannot identify the worst-case solution; in [143] the authors employ tools from variational analysis to get the (sub)-gradient formulae for probust functions with an uncertainty set.

Regarding energy management applications, we know that in a day-ahead market, we cannot react to the information revealed by the realizations of random vectors. In this case, when faced with optimization problems under uncertainty, we can make use of the static joint probabilistic constraints. However, for time-dependent energy problems, which take place in longer time periods, one should benefit again from the information flow acquired. In this case, one can make use of dynamic probabilistic constraints.

In chapter 4 we present the dynamic joint probabilistic constraints and prove novel structural results. The aim of our project was to generalize the results of structural analysis from the static probabilistic constraints to the dynamic one. This is important when decisions are time-dependent and when we can react on past observations. We studied the continuity and differentiability of the dynamic probability function. Strong and weak semicontinuity results are provided for the general case depending on whether policies are supposed to be in the L^p or $W^{1,p}$ spaces.

For a simple two-stage model, verifiable conditions for Lipschitz continuity and differentiability of this probability function are derived and endowed with explicit derivative formulae. This is of importance for an efficient solution of dynamic probabilistic constrained problems via numerical optimization.

Furthermore, we proved the non-convexity of the feasible set of decisions induced by a dynamic probability function in the L^p space.

In addition, we introduced the so-called Baker's dynamic problem and provided the necessary optimality conditions under a two-stage dynamic probabilistic function in L^2. For the two-stage hydro-power optimization problem (from the viewpoint of a social planner) in $W^{1,2}$ we were able to prove the existence of a solution. Finally, numerical results are illustrated for the solution of a two-stage hydro power model in L^2. In these numerical results, the second-stage optimal release policy is an affine function of the first-stage water inflow, for the linear objective of the producer's model and even for the non-linear objective of the social planner's model.

5.1 Future Work

The aim of future work is to further generalize the theoretical analysis of dynamic probabilistic constraints and to apply them in energy economic models.

Driven by structural and existence results for optimization problems with dynamic joint probabilistic constraints in $W^{1,2}$, we would like to further investigate necessary conditions of problem P_b and provide numerical solutions. In the context of hydro-power optimization, we already have explicit solutions for the two-stage hydropower dynamic probability functions. We would like to come up with KKT conditions for this problem (analogous to Theorem 4.36).

After generalizing the structural analysis from the static to the dynamic one, it would also be important to develop the stability analysis in the dynamic case. When formulating a stochastic optimization model, one tacitly assumes an underlying probability distribution to be given. In practical situations, however, this is rarely the case. One has to deal with incomplete information and approximations. This circumstance has motivated the stability analysis for optimal values and optimal solutions to stochastic programs with respect to perturbations of the underlying probability distributions, like in [115]. A way of treating the convergence of solutions for an optimization problem with static probabilistic constraints is done, for example, in [73].

After providing necessary and sufficient optimality conditions to problem P_b, one could analyze an economic question of importance: is it possible that weather uncertainty creates distinct opportunities for exercising market power under an oligopoly market structure? Market power is typically exercised by a reallocation of release of water between periods compared with what would be the socially desired release policy. Measuring the existence of market power by comparing price and marginal costs does not work for hydropower because variable cost is zero. The relevant variable cost is the opportunity cost of water; however, this is an expected variable and not directly observable. This problem is of importance and has been studied under the expected model in economic publications [67, 51].

Bibliography

[1] V. Acary, O. Bonnefon, and B. Brogliato. *Nonsmooth Modeling and Simulation for Switched Circuits*. Springer, 2011.

[2] W. Ackooij, R. Zorgati, Henrion R., and Möller A. Joint chance constrained programming for hydro reservoir management. *Optimization and Engineering*, 15:509–531, 2014.

[3] R. A. Adams and J. F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2003.

[4] D. Adelhütte, D. Aßmann, T. González Grandón, M. Gugat, H. Heitsch, R. Henrion, F. Liers, S. Nitsche, R. Schultz, M. Stingl, and D. Wintergerst. Joint model of probust constraints with application to gas network optimization. *Preprint of WIAS Berlin.*, (2550), 2018.

[5] L. Andrieu, R. Henrion, and Römisch W. A model for dynamic chance constraints in hydro power reservoir management. *European Journal of Operations Research*, 207:579–589, 2010.

[6] T. M. Apostol. *Calculus. Vol. I: One-variable calculus, with an introduction to linear algebra*. Second edition. Blaisdell Publishing Co. Ginn and Co., Waltham, Mass.-Toronto, Ont.-London, 1967.

[7] P. Artzner, F. Delbaen, J. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 2011.

[8] K. Atkinson and H. Weimin. *Spherical Harmonics and Approximations on the Unit Sphere: An Introduction*. Springer-Verlag Berlin, 2012.

[9] E. Beale, G. B. Dantzig, and R.D. Watson. A first order approach to a class of multi-time-period stochastic programming problems. *Mathematical Programming Study*, 27:103–117, 1986.

[10] A. Beck. *Introduction to Nonlinear Optimization: Theory, Algorithms and Applications*. SIAM, 2014.

[11] R. Bellman. The theory of dynamic programming. *Bull. Amer. Math. Soc.*, 60(6), 1954.

[12] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, 2009.

[13] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23:769–805, 1998.

[14] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25:1–13, 1999.

[15] A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88:411–424, 2000.

[16] D.P. Bertsekas. *Convexity, Duality, and Lagrange Multipliers*. Lecture Notes, Massaschussetts Institute of Technology, 2001.

[17] M. Bevers. A chance constraint estimation approach to optimizing resource management under uncertainty. *Canadian Journal of Forest Research*, 37:2270–2280, 2007.

[18] S. Bhattacharyya and Palit D. *Mini-Grids for Rural Electrification of Developing Countries*. Spinger Green Energy and Technology, 2014.

[19] J.R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer New York, 2010.

[20] V. I. Bogachev. *Measure theory. Vol. I, II*. Springer-Verlag, Berlin, 2007.

[21] J.F. Bonnans and A. Shapiro. Perturbation analysis of optimization problems. *Springer Series in Operations Research*, 2000.

[22] C. Borell. Convex set functions in d-space. *Periodica Mathematica Hungarica*, 6:111–136, 1975.

[23] M. Branda. Optimization approaches to multiplicative tariff of rates estimation in nonlife insurance. *Asia-Pacific Journal of Operations Research*, 31, 2014.

[24] H.J Brascamp and E.H. Lieb. On externsions of the brunn-minkowski and prékopa-leindler theorems, including inequalities for log-concave functions and with an application to the diffusion equations. *Journal of Functional Analysis*, 22:366–389, 1976.

[25] J. Brauchart, E. Saff, I. Sloan, and R. Womersley. QMC designs: optimal order Quasi Monte Carlo integration schemes on the sphere. *Mathematics of computation*, 83(290):2821–2851, 2014.

[26] J.S. Brauchart, E. Saff, I. Sloan, and R. Womersley. QMC designs: Optimal order Quasi Monte Carlo integration schemes on the sphere. *Mathematics of Computation*, 83:2821–2851, 2014.

[27] I. Bremer, R. Henrion, and A. Möller. Probabilistic constraints via sqp solver: Application to a renewable energy management problem. *Computational Management Science*, 260:3397–3447, 2015.

[28] O. Brezhneva and A. Tret'yakov. An elementary proof of the karush–kuhn–tucker theorem in normed linear spaces for problems with a finite number of inequality constraints. *Optimization*, 60:613–618, 2011.

[29] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, 2010.

[30] P. Carpentier, J. Chancelier, G. Cohen, Girardeau P., and M. De Lara. Dynamic consistency for stochastic optimal control problems. *Annals of Operations Research*, 200:247–263, 2012.

[31] A. Charnes and W. W. Cooper. Chance constrained programming. *Management Science*, pages 73–79, 1959.

[32] Cooper W. Charnes A. and G.H. Symmonds. Cost horizons and certainty equivalence: An approach to stochastic programming of heating oil. *Management Science*, pages 235–263, 1958.

[33] M. Chertkov, S. Misra, and M. Vuffray. Monotonicity of Dissipative Flow Networks Renders Robust Maximum Profit Problem Tractable: General Analysis and Application to Natural Gas Flows. *Mathematics: Optimization and Control*, 2015.

[34] G. Claßen, Koster A., Coudert D., and Nepomuceno N. Chance constrained optimization of reliable fixed broadband wireless networks. *INFORMS Journal on Computing*, 26:893–909, 2014.

[35] D. L. Cohn. *Measure theory*. Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]. Birkhäuser/Springer, New York, second edition, 2013.

[36] C.F. Colebrook. Turbulent Flow in Pipes with Particular reference to the Transition Region between the Smooth and Rough Pipe Laws. *J. Inst. of Civil Engineers*, I1:133–156, 1939.

[37] G. B. Dantzig. Linear programming under uncertainty. *Management Science*, (1):197–206, 1955.

[38] I. Deák. Three digit accurate multiple normal probabilities. *Numerische Mathematik*, 35:369–380, 1980.

[39] I. Deák. Computing probabilities of rectangles in case of multinormal distribution. *Journal of Statistical Computation and Simulation*, 26:101–114, 1986.

[40] I. Deák. Subroutines for computing normal probabilities of sets. *Annals of Operations Research*, 100:103–122, 2000.

[41] D. Dentcheva, A. Prékopa, and A. Ruszczyński. Concavity and efficient points for discrete distributions in stochastic programming. *Mathematical Programming*, 89:55–77, 2000.

[42] D. Dentcheva and A. Ruszczyński. Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, pages 548–566, 2003.

[43] J. Dupačová, Z. Gaivoronski, and T. Szantai. Stochastic Programming in water management: A case study and a comparison of solution techniques. *European Journal of Operations Research*, 52:28–44, 1991.

[44] H. Egger, T. Kugler, and N. Strogies. Parameter identification in a semilinear hyperbolic system. *IOPscience Inverse Problems*, 33, 2017.

[45] L. El Ghaoui and Lebret H. Robust solutions to least-squares problems with uncertain data. *SIAM J. Matrix Anal. Appl*, 18:1035–1064, 1997.

[46] J. Elstrodt. *Maß und Integrationstheorie*. Springer New York, 2011.

[47] R. Eustaquio, Karas E., and Ribeiro A. Constraint qualifications for nonlinear programming. *Encuentro Nacional de Docentes en Investigación Operative (ENDIO)*, 2008.

[48] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.

[49] M. H. Farshbaf-Shaker, R. Henrion, and D. Hömberg. Properties of chance constraints in infinite dimensions with an application to pde constrained optimization. *Set-Valued and Variational Analysis*, Oct 2017.

[50] W. H. Fleming and H. M. Soner. *Controlled Markov processes and viscosity solutions*, volume 25 of *Stochastic Modelling and Applied Probability*. Spinger New York, 2006.

[51] F. R. Førsund. *Hydropower Economics*, volume 217. Springer, 2015.

[52] Yeh W. G. Reservoir management and operations models: A state-of-the-art review. *Water Resources Research*, 21:1797–1818, 1985.

[53] J. Garnier, A. Omrane, and Y. Rouchdy. Asymptotic formulas for the derivatives of probability functions and their monte carlo estimations. *European Journal of Operations Research*, 198:848–858, 2009.

[54] W. Garvin, H. Crandall, and Spellmann R. Applications of linear programming in the oil industry. *Management Science*, 3:407–430, 1957.

[55] A Genz. Numerical computation of multivariate normal probabilities. *J. Comp. Graph Stat.*, 1:141–149, 1992.

[56] A Genz and F. Bretz. *Computation of Multivariate Normal and t Probabilities*. Spinger-Verlag Berlin, 2009.

[57] P.E. Gill, W. Murray, and M.H. Wright. *Practical Optimization*. Academic Press, 1981.

[58] J. Goh and M. Sim. Distributionally robust optimization and its tractable approximations. *Operations Research*, 58:902–917, 2010.

[59] A. S. Goldberger. *A course in Econometrics*. Harvard University Press, 1991.

[60] T. González Grandón, H. Heitsch, and R. Henrion. A joint model of probabilistic/robust constraints for gas transport management in stationary networks. *Computational Management Science*, 14(3):443–460, 2017.

[61] T. González Grandón, R. Henrion, and P. Pérez-Aros. Dynamic probabilistic constraints under continuous random distributions. *Preprint*

of the Collaborative Research Center TRR 154. https://opus4.kobv.de/opus4-trr154/frontdoor/index/index/searchtype/latest/docId/254/start/0/rows/10, (254), 2019.

[62] C. Gotzes, H. Heitsch, R. Henrion, and R. Schultz. Feasibility of nominations in stationary gas networks with random load. *Mathematical Methods of Operations Research*, 84:427–457, 2016.

[63] C. Gotzes, S. Nitsche, and R. Schultz. Probability of feasible loads in passive gas networks with up to three cycles. Preprint, available online at OPUS TRR154, 2017.

[64] N. Gröwe, W. Römisch, and R. Schultz. A simple recourse model for power dispatch under uncertain demand. *Annals of Operations Research*, 59:135–164, 1995.

[65] V. Guiges and R. Henrion. Joint dynamic probabilistic constraints with projected linear decision rules. *Optimization Methods and Software*, 2016.

[66] M. Guignard. Generalized kuhn-tucker conditions for mathematical programming problems in banach soace. *SIAM Journal on Control*, 7:232–241, 1969.

[67] P. V. Hansen. Inflow uncertainty in hydropower markets. *The Scandinavian Journal of Economics*, 11(1):189–207, 2009.

[68] J.P. Hennessey. Some aspects of wind power statistics. *J Applied Meteorology*, 16:119–128, 1977.

[69] R. Henrion. Structural properties of linear probabilistic constraints. *Optimization: A Journal of Mathematical Programming and Operations Research*, 56:425–440, 2007.

[70] R. Henrion. Optimierungsprobleme mit wahrscheinlichkeitsrestriktionen: Modelle, struktur, numerik. Lecture Notes, 2010.

[71] R. Henrion. A gradient formula for linear chance constraints under gaussian distribution. *Mathematics of Operations Research*, 2:437–488, 2012.

[72] R. Henrion and A. Möller. A gradient formula for linear chance constraints under Gaussian distribution. *Mathematics of Operations Research*, 37:475–488, 2012.

[73] R. Henrion and W. Römisch. Hölder and Lipschitz stability of solution sets in programs with probabilistic constraints. *Mathematical Programming*, 100:589–611, 2004.

[74] R. Henrion and W. Römisch. Lipschity and differentiability properties of quasiconcave and singunal normal distribution functions. *Annals of Operations Research*, 177:115–125, 2010.

[75] R. Henrion and C. Strugarek. Convexity of chance constraints with independent random variables. *Computational Optimization and Applications*, 41:263–276, 2008.

[76] A. D. Ioffe. Necessary and sufficient conditions for a local minimum 3: Second order conditions and augmented duality. *SIAM J. Control Optim.*, 17:266–288, 1979.

[77] P. Kall and J. Mayer. *Stochastic Linear Programming: Models, Theory and Computation*. Springer, 2011.

[78] W. Karush. Minima of functions of several variables with inequalities as side conditions. *Msc. Thesis University of Chicago*, 1939.

[79] D.S. Kirschen and G. Strbac. *Fundamentals of Power System Economics*. Wiley, 2004.

[80] W. Klein Haneveld, M. Streutket, and M. van der Vlerk. An alm model for pension funds using integrated chance constraints. *Annals of Operations Research*, 177:47–62, 2010.

[81] T. Koch, B. Hiller, M. Pfetsch, and L. Schewe. *Evaluating gas network capacities*, volume 21 of *MOS-SIAM Series on Optimization*. SIAM Philadelphia, 2015.

[82] H. Kuhn and A. Tucker. Nonlinear programming. *Proceedings of the Second Berkley Symposium on Mathematical Statistics*, pages 481–492, 1951.

[83] G. Leobacher and Pillichshammer F. *Introduction to Quasi-Monte Carlo Integration and Applications*. Birkhäuser Verlag, 2014.

[84] Z. Li and C. Floudas. A comparative theoretical and computational study on robust counterpart optimization: Probabilistic guarantess on constraint satisfaction. *Ind. Eng. Chem.*, 2011.

[85] D. G. Luenberger. *Optimization by Vector Space Methods*. Wiley, 2016.

[86] Glavic M., M. Hajian, Rosehart W., and Zareipour H. A Chance Constraint Optimization Approach for Control of Transmission Voltages. *IEEE Transactions on Power Systems*, 27, 2012.

[87] F. Maggioni, Potra F., and Bertocchi M. A scenario-based framework for supply planning under uncertainty: stochastic programming versus robust optimization approaches. *Computational Management Science*, 14:5, 2017.

[88] O.L. Mangasarian and S. Fromovitz. The fritz john necessary optimality conditions in the presence of equality and inequality constraints. *Math. Anal. and Appl.*, 17:37–47, 1967.

[89] A. Manne. A linear programming model of the us petroleum industry. *Econometrica*, 26(67-106), 1972.

[90] K. Marti. Differentiation of probability functions: The transformation method. *Math. Programming*, 75:201–220, 1996.

[91] B.L. Miller and H.M. Wagner. Chance constrained programming with joint constraints. *Operations Research*, 13:930–945, 1965.

[92] J. Monahan and Genz. A. Spherical-radial integration rules for bayesian computation. *Journal of the American Statistical Association*, 92:664–674, 1997.

[93] B. S. Murdokhovich and N. Mau Nam. *An Easy Path to Convex Analysis and Applications*. Morgan and Claypool Publishers, 2014.

[94] H. Niederreiter. Random number generation and quasi-monte carlo methods. *SIAM*, 1992.

[95] J. Nocedal and Wright S. *Numerical Optimization*. Springer New York, 1999.

[96] N. J. Olieman and van Putten B. Estimation method of multivariate exponential probabilities based on simplex coordinates transform. *J. Stat. Comput. Simul.*, 80:355–361, 2010.

[97] D. Peterson. A review of constraint qualifications in finite-dimensional spaces. *SIAM Review*, 15(3), 1973.

[98] J. Peypouquet. *Convex Optimization in Normed Spaces: Theory, Methods and Examples*. Springer Briefs in Optimization, 2015.

[99] M. Pfetch, A. Fügenschuh, B. Geißler, N. Geißler, R. Gollmer, B. Hiller, J. Humpola, R. Henrion, T. Koch, T. Lehmann, M. Schmidt, R. Schultz, R. Schwarz, J. Schweiger, C. Stangl, M. Steinbach, and B. Willert. Mathematical optimization for challenging network planning problems in unbundled liberalized gas markets. *Energy Systems*, 5:449–473, 2014.

[100] M. Pfetch, A. Fügenschuh, B. Geißler, N. Geißler, R. Gollmer, B. Hiller, J. Humpola, T. Koch, T. Lehmann, A. Martin, R. Schultz, R. Schwarz, J. Schweiger, C. Stangl, M. Steinbach, and B. Willert. Validation of nominations in gas network optimization: models, methods, and solutions. *Optimization Methods and Software*, 30(1):15–53, 2015.

[101] B. Polyak. *The Introduction to Optimization*. Optimization Software, New York, 1987.

[102] A. Prékopa. On Probabilistic Constrained Programming. *Proceedings of Princeton Symposium on Mathematical Programming*, 1970.

[103] A. Prékopa. Contributions to the theory of stochastic programming. *Mathematical Programming*, 4:202–221, 1973a.

[104] A. Prékopa. On logarithmic concave measures and functions. *Acta Science Mathematics*, 34:335–343, 1973b.

[105] A. Prékopa. Logarithmic concave measures and related topics. *Stochastic Programming*, 1980.

[106] A. Prékopa. Programming under Probabilistic Constraint and Maximising a Probability under Constraints. *Rutcor Research Report*, pages 35–93, 1993.

[107] A. Prékopa. *Stochastic programming*, volume 324 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1995.

[108] A. Prékopa. On the Concavity of Multivariate Probability Distribution Functions. *Operations Research Letters*, 29:1–4, 2001.

[109] A. Prékopa and T. Szántai. A new multivariate gamma distribution and its fitting to empirical streamflow data. *Water Resources Research*, 14:19–24, 1978.

[110] L. Prince, B. Purrington, and Ramsey J. Gasoline blending at texaco using nonlinear programming. *Joint National Meeting Chicago TIMS/ORSA*, 1983.

[111] E. Raik. Qualitative research into the stochastic nonlinear programming problems. *Eesti NSV Teaduse Academia Toimetised*, 20:8–14, 1971.

[112] K. S. Rangarajan. *A First Course in Optimization Theory*. Cambridge University Press, 1996.

[113] W.D. Richter. A geometric approach to the gaussian law. *Symposia Gaussiana, Walter de Gruyter*, pages 25–46, 1994.

[114] S. M. Robinson. Stability theory for systems of inequalities, part ii: Differentiable nonlinear systems. *SIAM Journal on Control and Optimization*, 38:497–513, 2000.

[115] W. Römisch and R. Schultz. Stability analysis for stochastic programs. *Annals of Operations Research*, 30:241–266, 1991.

[116] W. Römisch and R. Schultz. Decomposition of a multi-stage stochastic porgram for power dispatch. *ZAMM-Zeitschrift f"ur Angewandte Mathematik und Mechanik*, 76:29–32, 1996.

[117] W. C. Rooney and L. T. Biegler. Optimal process design with model parameter uncertainty and process variability. *AIChE J.*, 49:483–449, 2003.

[118] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, 1987.

[119] Walter Rudin. *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, second edition, 1991.

[120] A. Ruszczynski and A. Shapiro. *Stochastic Programming*. Handbooks in Operations Research and Management Science, 2003.

[121] A. Saumard and Wellner J. A. Log-concavity and strong log-concavity: a review. *Statistics surveys*, 2014.

[122] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming*, volume 9 of *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, second edition, 2014. Modeling and theory.

[123] M. Solodov. Constraint qualifications. *Encyclopedia of Operations Research and Management Science*, 2011.

[124] A. L. Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21:1154–1157, 1973.

[125] J. P. Stern and H. Rogers. The dynamics of a liberalised european gas market – key determinants of hub prices, and roles and risks of major players, 2014.

[126] G. Stock and M. Henle. Integration virtueller kraftwerke in querverbunsysteme. *Euroheat and Power, Fernwärme international*, 31:58–63, 2002.

[127] G. Symonds. Linear programming. the solution of refinery problems. *Esso Standard Oil Company*, 1955.

[128] T. Szantai. Numerical evaluation of probabilities concerning multidimensional probability distributions. PhD. Thesis Hungarian Academy of Sciences, 1985.

[129] T. Szántai. Improved bounds and simulation procedures on the value of the multivariate normal probability distribution function. *Annals of Operations Research*, 100:85–101, 2000.

[130] L. Tonelli. Sur un méthode directe du calcul des variations. *Rend. Circ Mat. Palermo*, 39:233–264, 1915.

[131] European Union. Directive 98/30/ec of the European Parliament and of the Council concerning common rules for the internal market in natural gas. *European Comission*, 204:0001:0012, 1998.

[132] European Union. Directive 2003/55/ec of the European Parliament and of the Council concerning common rules for the internal market in natural gas and repealing directive 98/30/ec. *European Comission*, 2003.

[133] European Union. Directive 2009/73 of the European Parliament and of the Council concerning common rules for the internal market in natural gas. *European Comission*, 211:0094:0136, 2009.

[134] S. Uryasev. A differentiation formula for integrals over sets given by inclusion. *Numerical functional Analysis and Optimization*, 10:827–841, 1989.

[135] S. Uryasev. Derivatives of probability functions and integrals over sets given by inequalities. *Journal of Computational and Applied Mathematics*, 56:197–223, 1994.

[136] S. Uryasev. Derivatives of probability functions and some applicatons. *Annals of Operations Research*, 56:287–311, 1995.

[137] W. van Ackooij. *Chance Constrained Programming: with applications in Energy Management.* PhD thesis, Ecole Centrale Paris, 2013.

[138] W. van Ackooij. A comparison of four approaches from stochastic programming for large-scale unit commitment. *Euro J. Computational Optimization*, 2015.

[139] W. van Ackooij, A. Frangioni, and W. de Oliveira. Inexact stabilized Benders' decomposition approaches with application to chance-constrained problems withfinite support. *Computational Mathematics and Applications*, 65:637–669, 2016.

[140] W. van Ackooij and R. Henrion. Gradient formulae for nonlinear probabilistic constraints with Gaussian and Gaussian-like distributions. *SIAM J. Optim.*, 24(4):1864–1889, 2014.

[141] W. van Ackooij and R. Henrion. (sub-) gradient formulae for probability functions of random inequality systems under gaussian distribution. *J. Uncertainty Quantification*, 5:63–87, 2017.

[142] W. van Ackooij, R. Henrion, A. Möller, and R. Zorgati. On joint probabilistic constraints with gaussian coefficient matrix. *Operations Research Letters*, 39:99–102, 2011.

[143] W. van Ackooij, R. Henrion, and P. Pérez-Aros. Generalized gradients for probabilistic/robust (probust) constraints. *WIAS Preprint no. 2569*, 2019.

[144] W. van Ackooij, Henrion R., Möller A., and R. Zorgati. On probabilistic constraints induced by rectangular sets and multivariate normal distributions. *Mathematical Methods of Operations Research*, 71:535–549, 2010.

[145] W. van Ackooij and C. Sagastizábal. Constrained bundle methods for upper inexact oracles with applications to joint chance constrainted energy problems. *SIAM Journal on Optimization*, 24:733–765, 2014.

[146] H. J. Wagner and J. Mathur. *Introduction to Hydro Energy Systems: Basics, Technology and Operation.* Springer-Verlag Berlin, 2011.

[147] P. Wong and R. Larson. Optimization of natural-gas pipeline systems via dynamic programming. *IEEE Transactions on Automatic Control*, 13(5), 1968.

[148] J. Yong and Zhou X. *Stochastic Controls, Hamiltonian systems and HJB equations*. Springer Verlag, New York, 2000.

[149] W. Zwangwill. *Nonlinear Programming. A Unified Approach*. Prentice-Hall, Englewood Cliffs, 1970.